PSYCHOMETRIC VALIDATION OF THE HISPANIC BILINGUAL GIFTED SCREENING INSTRUMENT: AN ITEM RESPONSE THEORY APPROACH

A Dissertation

by

YINAN FAN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Rafael Lara-Alecio
Co-Chair of Committee,	Fuhui Tong
Committee Members,	Myeongsun Yoon
	Beverly J. Irby
	Yeping Li
Head of Department,	Victor Willson

August 2014

Major Subject: Educational Psychology

ABSTRACT

Demographics in the United States continue to shift with a rapidly growing Hispanic population. On the other hand, a mismatch still exists between Hispanic students' enrollment in gifted and talented (G/T) programs and general programs. The under-representation of Hispanic students in G/T programs has been attributed to a lack of proper instrument to identify gifted students who are linguistically and culturally diverse; insufficient preparation of teacher in the initial teacher referral phases; and ambiguous definitions of intelligence and giftedness.

In this study I investigated psychometric properties of the Hispanic Bilingual Gifted Screening Instrument (HBGSI) within an item response theory (IRT) framework. The HBGSI was developed with social-cultural context in mind and has been recommended for use in the first phase of teacher referral process. Participants in this study were Hispanic bilingual students in first to third grade, who participated in a large-scale longitudinal randomized study carried out in a Texas urban school district. The purpose of this study was to further validate HBGSI within the framework of IRT, exploring the factor structure and dimensionality of the instrument at the item level. I further tested the possibility of constructing an abbreviated version of HBGSI with fewer items for ease of administration, which would potentially lower the demand on the teacher's time, enhance accessibility and facilitate increased usage of the instrument.

Results revealed a bifactor structure with a strong general factor corresponding to overall giftedness among Hispanic bilingual students, and five domain factors regarding social responsibility, academic achievement, creative performance, problem solving, and native language proficiency. The multidimensional bifactor IRT model provided information related to each item concerning its discriminating power, thresholds and information regarding the latent constructs. Best items were selected while preserving the integrity of the original HBGSI, and cutting the length to almost half. Thus an abbreviated version of HBGSI was feasible and the adaptation is presented. Overall, this study further validated that the HBGSI holds promise in screening potential Hispanic bilingual students in elementary grades.

DEDICATION

To my beloved family

And to the great State of Texas: them trees, grass, and wildflowers; them cows, dogs, and horses; them cowgirls and cowboys, the people, and the nothingness.

ACKNOWLEDGEMENTS

My scholarly work would not be possible without help from many individuals. My genuine appreciations go to my co-chairs Dr. Rafael Lara-Alecio and Dr. Fuhui Tong, who have been most supportive to me in many ways throughout my academic career. They were there for me every step of the way, from selecting courses, to forming dissertation topics, to the completion of the dissertation. They also graciously involved me in several national and international research projects and such research experiences have been invaluable for my growth as a researcher. Dr. Lara and Dr. Tong taught me more than academics. In many talks with Dr. Lara, I learned his vision for education, the bumpy but rewarding road to success, and many aspects of the Hispanic population. In late-night and weekend emails with Dr. Tong, I leaned the spirit of hardwork and perseverance. I am grateful to have them as my advisors and committee chairs.

I am also thankful to Dr. Beverly J. Irby, who is such a bright star, but stayed approachable to me. As one of the authors of HBGSI, Dr. Irby has been most generous in letting me work with the instrument and in always sharing crucial insights every time I discussed the instrument. She helped me in closely reviewing conference presentations, reviewing drafts, and going over each item in HBGSI and paragraph. Beyond HBGSI, Dr. Irby has also opened doors for me in many ways in conferences and publications, to which I am greatly thankful.

I would also like to thank Dr. Myeongsun Yoon. It is in her classes of advanced psychometrics and IRT that I started to gain understanding and growing interest related to my dissertation topic. She is an expert in IRT and statistical modeling, and her key

advice has been critical to me, as I approached the topic in various ways, with sometimes failed attempts. She has always been very patient with me in the various phases of my study.

Dr. Yeping Li is another person to whom I am obliged. Despite his busy schedules, he was always available to talk to, has tolerated my mindless chats, and been as well, very patient with me. Dr. Li has also welcomed me in many TLAC presentations where important figures were invited to talks about their current research, which greatly broadened my horizon and nourished my mind. But still, I will always remember wise words from Dr. Li, to work hard on real things, to have a goal worthwhile.

So many other people made my graduate life memorable. I am fortunate to have met such great faculty members, staff members, and fellow graduate students. A full list is beyond the scope here, but thank you for indulging and endowing me with knowledge, for making paperwork much easier than paper, and for the time when we shared things in academic life, and life not accounted by academia. To my dearest friends, thank you for staying true to me even though we lose touch some times.

Last but definitely not least, I want to thank my family, who gave me unconditional support throughout the pursuit to my dream, who believed in me even at times when I lose sight of the future, and who surrounded me with so much love that made life meaningful. I look forward to share the next chapters in my life with all of you.

OK, lastly, if you are from the Amazing Race and are actually reading this, thank you for the great show, and hope to see you in a race！

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION AND BACKGROUND

The changing demographics of America are shaping new landscape in America's public school system. U.S. Census (2010) data showed that from 2000 to 2010, more than half of the nation's total population growth is composed mostly of those who self-reported as Hispanic/Latino. More than 50 million people (or 16% of the total population) of Hispanic or Latino origin resided in the United States in 2010, an increase of over 15 million in sheer number and about 3% of the total population from 2000. In three states resides the highest number of the Hispanic population: California, Texas and Florida; while percentage wise, New Mexico, Texas, and California takes the top three . In Texas (where this study is conducted) alone, 34.7% of the population, 5 years and older, speak languages other than English at home according to 2011 American Community Survey (Ryan, 2013). For children growing up in families that speak another language at home, school means more than just learning academic contents, but also learning a new language of English, especially English as academic language.

Enrollment in public schools in Texas well reflected the changes of the ethnic diversity. In the academic year of 2012-2013, Hispanic students are not only the largest ethnic group enrolled in Texas public schools, but in fact outnumbered the total of all other ethnic groups added together. According to data from Texas Education Agency ([TEA], 2013), the proportion of public school enrollment accounted for by Hispanic students went from 39.6% in 1999 to 51.3% in 2013. In the school year of 2012- 2013, 51.3% (2,606,126) out of a total of 5,075,840 students served in Texas public schools were reported as Hispanic students (TEA, 2014). The percentage is higher in earlier grades than higher grades. Almost one sixth (17.0%) of all public school students were receiving instructional services in bilingual or English as a second language (ESL), and an overwhelming majority (90.2%) of these students were Hispanic students (TEA, 2014). These Hispanic students constitute the largest group of English language learners (ELLs) and have unique academic needs due to their linguistic and cultural diversity.

Underrepresentation in G/T Programs

While public school system personnel are making every effort to respond to the needs of their changing students, mismatches still exist in every level of those systems. In the realm of gifted education, there is an underrepresentation of Hispanic children in gifted and talented (G/T) programs in comparison to the general enrollment data aforementioned. In the same academic year of 2012- 2013, in G/T programs, Hispanic students accounted for only 40.0% of the overall enrollment, much less than its overall student representation percentage (TEA, 2014). Evidently, “discrepancies still exist between the percentages of underrepresented populations in the total student population versus the percentage of underrepresented populations identified for G/T services” (TEA, 2008, para. 2).

One of the reasons that Hispanic bilingual students have traditionally been under-represented in gifted education programs is a lack of instruments that can properly identify such students: the identification and assessment process is difficult because it involves students with cultural and linguistic background different from the native English-speaking children from middle-class families (Cohen, 1998). When students whose native language is not English take tests in English, their language proficiency becomes an additional construct that influences performance, besides the target construct that the test is designed to measure.

Ambiguous definitions of giftedness also contributed to the problem (Irby, Lara-Alecio & Rodriguz, 2003b; McKenzie, 1986). Giftedness has very often been operationally equated to scores on an intelligence test. These standardized tests themselves, not only encompass a likely too narrow definition of giftedness, but also have been found to be biased against minority students (e.g., Gonzalez & Yawkey, 1993).

There is consequently a lack of teacher preparation and understanding of linguistically and culturally diverse (LCD) students, as teachers were not prepared from their training to properly identify giftedness in LCD students, to pay attention to the underachievement of G/T minority students and to be sensitive about their social and emotional needs (Ford & Graham, 2003). This in turn, reinforces the under-representation of LCD students as they are likely to be excluded for referral in G/T programs.

Therefore it is important that effective and proper ways are established to better identify Hispanic students that are gifted and talented. One instrument, the Hispanic Bilingual Gifted Screening Instrument (HBGSI), was developed with such purpose in mind. It was created to be used in the first phase of GT identification, specifically for the group of Hispanic bilingual children (Irby & Lara-Alecio, 1996, Irby, Lara-Alecio, & Rodriguez, 2003a). Its psychometric properties have been explored in several previous studies (e.g., Contreras-Vanegas, 2011; Esquierdo, 2006; Fultz, 2004;), and the instrument has been shown to be effective in teacher referral process (Irby, Hernandez, Torres, & Gonzalez, 1997, as cited in Fultz, 2006).

Definition of Terms

Hispanic

The term is derived from the Latin word *Hispania* and is used to describe people who trace their origins to Spain and the Spanish-speaking countries of Latin America (Castellano, 2011, p. 256).

English Language Learners

English language learners are those who are beginning to learn English or who have not demonstrated proficiency in English (Padron &Waxman, 1999).

Measurement

A measurement model relates performance on the behavior sample (i.e. test) to the latent variable (Embretson, 1999).

Latent Variable

A variable that is not directly observed but inferred from observed variables. Unobservable quantities such as errors are not usually described as latent.

Validity

According to psychometric standards jointly published by American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) in 1999, validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests. Validity is the most fundamental consideration in developing and evaluating tests. Professional judgment guides decisions regarding the specific forms of evidence that can best support the intended interpretation and use.

Reliability

Reliability refers to the consistency of measurements when the testing procedure is repeated on a population of individuals or groups. However, no single examinee is completely consistent, and in some instances, because of subjectivity in the scoring process, an individual's obtained score and the average score of a group will always reflect at least a small amount of measurement error. Information about measurement error is essential to the proper evaluation and use of a test instrument (AERA, APA, & NCME, 1999).

Test Development

Test development is the process of producing a measure of some aspect of an individual's knowledge, skill, ability, interests, attitudes, or other characteristics by developing items and combining them to form a test, according to a specified plan. Test development also includes specifying conditions for administering the test, determining procedures for scoring the test performance, and reporting the scores to test users (AERA, APA, & NCME, 1999).

Test Revisions

Tests and their supporting documents are reviewed periodically to determine whether revisions are needed. Revisions and amendments are necessary when new research data, significant changes in the domain, or new conditions of test use and interpretation would either improve the validity of interpretation of the test scores, or suggest that the test is no longer fully appropriate for its intended use. Revisions to test content are also made to ensure the confidentiality of the test (AERA, APA, & NCME, 1999).

Classical Test Theory

Classical test theory (CTT) model is based on several assumptions about error distributions: errors have an expected value of zero, are normally and uniformly distributed in persons, and are uncorrelated with all other variables. Classical Test Theory is limited in several ways.

Item Response Theory

Item response theory (IRT) is the set of measurement principles that evaluates how well tests and individual items in a test work. It can be used for testing development and has made adaptive testing possible (Embretson & Hershberger, 1999).

Statement of Problem

Giftedness can be found in all racial and ethnic groups, and all social classes. However, in today's public education system, there is still a discrepancy between Hispanic representation in G/T programs and in general enrollment (TEA, 2008; TEA, 2014). The significant underrepresentation of linguistically and culturally diverse (LCD) students in G/T programs has caused concerns in academia and in the broad society (e.g, Ford, 1998; National Reserch Council, [NRC], 2002).

Many researchers attributed the underrepresentation to an absence of adequate assessment procedures and programming for gifted LCD students (e.g., Bermúdez and

Rakow, 1990; Castellano, 1995; Ford, 1995, 1998). Assessment tools and procedures for gifted programs that traditionally relied heavily on English oral and written language skills (e.g., Hartley, 1987) clearly put ELLs at a disadvantage. Narrow conception of giftedness that equates giftedness to academic achievement and intelligence reflected by traditional IQ scores limited chance of identification for linguistically and culturally diversed (LCD) students (Ford & Grantham, 2003). Besides content of tests, negative social views against minority students possibly triggered stereotypical threat and results in under estimation of grades and test scores (Walton & Spencer, 2009). Misconceptions, or lack of understanding of cultural diversity played against minority students too. Cultural values are reflected in students' behaviors, but those different from the mainstream can be misinterpreted by the majority culture (Bermúdez, Rakow, Márquez, Sawyer, & Ryan, 1991). Serious inequity results if schools undervalue behaviors that certain cultures foster (Guild, n.d.). For example, in many cultures, it is a virtue to listen, learn and form mature thoughts before one speaks, but in another culture, such behavior might be interpreted as slowness (Hartley, 1987).

Sternberg (1985) has argued that intelligence, or giftedness is developed within each cultural context, but many assessments only judge through the lens of mainstream culture and values. Bennet (1986) warned us that if classroom expectations are limited by one cultural orientation, success of learners guided by another cultural orientation will be impeded.

It is important that a comprehensive understanding of intelligence and giftedness is promoted in schools. However, teachers have been found to have inadequate preparation for gifted education in a multicultural setting (Ford, 1999). Teacher rating scaled were advocated as part of the identification process, as it involves teacher input (Renzulli, Hartman, & Callahan, 1971) and can be used for assessing students social skills (e.g. Gresham & Elliott, 1990), adjustment behaviors (McDermott, Marston, & Stott, 1993), and in addition, help teachers understand and better identify students in their linguistic and cultural backgrounds. If initial screening instruments that include appropriate, referential, operational definitions and characteristics of giftedness Hispanic, bilingual students are not in place, then these students will continue to be denied access to programs due to their inability to move beyond the screening phase (Irby & Lara-Alecio & Rodriguez, 2003). If gifted or potentially gifted children were not properly identified and developed due to their diversity background, it will be the loss of our society in the future. Psychometric properties in these teacher ratings scales are important aspects of identification process, as they provide structured ways for teachers to gain understanding across a variety of constructs associated with outstanding achievement (Erwin &Worrell, 2012).

Significance of Study

In order to identify giftedness and potential giftedness, criteria using multiple sources of assessments that takes into context of the student, school and context has been suggested (e.g., NRC, 2002; National Association of Gifted Children, [NAGC], 2009; Worrel, 2009). As part of this effort, HBGSI was developed as a screening instrument for the first phase to identify gifted and potentially gifted students in the Hispanic bilingual group (Irby & Lara-Alecio, 1996; Irby, Lara-Alecio, & Rodriguez, 2003a).

Using teacher ratings, HBGSI studied students from multiple perspectives including linguistic, cultural, cognitive, familial and social characteristics, and detected giftedness and potential giftedness from a student's typical behaviors in their everyday life.

Validation evidence for HBGSI has been provided in previous studies, however, no such studies have been conducted at item-level. However, to date, no validation studies of gifted screening instruments for minority students were found, that was conducted using Item Response Theory (IRT) for validation and development. In this study, I would like to explore properties of each item, investigate dimensionality of HBGSI, and provide validation evidence within the measurement framework of Item Response Theory (IRT). Based on the item level findings, it is likely that a shortened version of HBGSI can be recommended. Besides psychometric validation, the practical benefits include that teachers can spend less time on this instrument when assessments for each student in classroom are needed. The instrument also helps teacher gain more insights to characteristics of Hispanic Bilingual gifted and talented students, so that these students can not only be referred to G/T programs, but also develop their gifted potentials with the guidance of their teachers.

Purpose of Study

The purpose of this study was to further validate HBGSI within the framework of IRT, exploring the factor structure and dimensionality of the instrument at the item level. I further tested the possibility of constructing an abbreviated version of HBGSI with fewer items for ease of administration, which would potentially lower the demand on the teacher's time, enhance accessibility and facilitate increased usage of the instrument.

Research Questions

This study was guided by the following research questions:

1. What is the dimensionality of HBGSI, and what is the degree of saturation by a general factor?
2. Based on the IRT model, what are the properties of each item?
3. Based on the IRT model, is it possible that a shortened version of HBGSI can be recommended, and what items will it include?

Overview of Study

This study is presented in five chapters. Chapter I provides background of the study, statement of the problem, purpose and significance of study, as well as specific research questions. The second chapter elaborates on literature review that concerns three aspects of the problem: definition of giftedness; under-representation of Hispanic students in G/T programs; and the framework of Item Response Theory and its applications.

Methodology is presented in Chapter III, including study context, participants, instrumentation, data collection and approaches to analysis. In the subsequent chapter, results are presented answering each of the research questions in order. Based on these results, final discussion and conclusion were drawn in Chapter V along with implications for practices and recommendations.

CHAPTER II

LITERATURE REVIEW

In this chapter, literature concerning giftedness, its definition and its identification, especially in the Hispanic bilingual population is reviewed. Multiple viewpoints are taken in order to examine what qualities possessed by children are considered characteristics of giftedness.

These issues were further related to the under-representation of Hispanic bilingual students in G/T programs. In the end, item response theory is introduced in a nutshell, for readers who might not be familiar with this method.

Giftedness and Intelligence

To identify giftedness, it is essential to understand what giftedness is and how it is manifested. However, a clear definition of giftedness is not available in literature, although it is traditionally associated with the higher end of intelligence. While intellectual pursuit has been at work driving progress in scientific discoveries, technological innovations, and social reforms, a vast amount of discussions also accumulated in the quest for a clear definition of intelligence itself. This discussion is still ongoing in academia without consensus.

Intelligence Tests

One of the most influential contributions in understanding intelligence was made by French psychologist Alfred Binet. Commissioned by the French government to identify students who would less likely succeed in public schools at the time and might need alternative education, Binet and Simon created a scale that comprised of tasks of increasing difficulties, later known as Binet-Simon scale (Binet & Simon, 1916). Their choice of style was to make all tests "simple, rapid, convenient, precise, heterogeneous, holding the subject in continued contact with the experimenter" (Binet & Simon, 1916, p. 41). Judgment, interpreted as "good sense, practical sense, initiative, the faculty of adapting one's self to circumstances" was taken as the most important factor for intelligence by Binet, while tests of memories were utilized in order to gain "an appreciation of judgment" (Binet & Simon, 1916, p. 41). The assumption of a single faculty of intelligence was made in Binet-Simon scale is different from a lot of other practices at the time, which attempted measure as separate functions, memory, attention, sense discrimination, etc.

Mental age was reported as a standard in Binet-Simon scale, instead of vague and non-uniform descriptive adjectives commonly used with clinical physicians at the time. The use of numbers enabled quantitative manipulation. By dividing one's mental age with actual age, the term intelligence quotient (IQ) were then used to represent the level of intelligence he/she processed (Binet & Simon, 1916).

The Binet-Simon scale greatly influenced the understanding of intelligence in the United States. A strong advocate was Lewis Terman (Terman, 1916) who published an adapted scale in the book entitled *Measurement of Intelligence*, even though Binet himself seemed cautious in using the word 'measurement' but preferred the term 'classification' (Binet & Simon, 1916, p. 44). The revised scale, published with newly added items and re-ordered original items and was normed to an American sample, was later known as the Stanford-Binet Intelligence Scale (Becker, 2003).

Following the trend, several other forms of IQ tests were developed and administered to a wide population. IQ scores have become the standard to measure human intelligence. In 1917 at the outset of World War I (WWI), two tests were developed by psychologist Robert Yerkes. They were administered to over two million soldiers to help the army recruit and select those who were suited for leadership and specific positions (McGuire, 1994). These tests remained to be used in a wide variety of situations after WWI, not without critique in today's point of views.

Another form of popular test was the Wechsler series, such as the Wechsler-Bellevue Intelligence Scale (Wechesler, 1939) for adults; its revised versions of Wechsler Adult Intelligence test (WAIS; Wechesler, 1955, 1997, 2008); and the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949, 1974, 1991, 2003). Point scale was used instead of age scale, so partial credits were taken into account. Multiple subtests with different content areas replaced one single score to provide a portfolio of point scores that can reported both overall and respective sub-area abilities. A non-verbal performance scale was utilized too.

Subtest structures have also been adopted by later versions of Stanford-Binet scale, which remained one of the most popular assessment tools used (Roid & Barram, 2004). The latest version, Stanford-Binet Intelligence Scale, Fifth Edition (SB5) provides comprehensive coverage of five factors of cognitive ability: Fluid Reasoning, Knowledge, Quantitative Processing, Visual-Spatial Processing, Working Memory (Stanford-Binet, 2003).

The g-factor

The idea of general intelligence for all human abilities, as assumed by Binet and Terman, was first hypothesized by Galton (1869), while empirical evidence was provided by Spearman (1904, 1927), who studied a number of mental tests available at the time and introduced the method of factor analysis. Spearman's findings revealed one common source of variance that explained the correlations between different measurements, that he named *g*. The psychometric construct *g*, or Spearman's *g*, however, was considered a narrower conception of general ability than Galton's notion, conceived with both biological and evolutionary terms (Jenson, 1986).

Multifaceted scales, with their division of a scale into multiple subtests did not prevent a common factor from emerging. Hierarchical factor analyses were used to explore relationships between subtests, and a general factor at the highest level was found. For example, a hierarchical factor analysis of the 12 subtests of WISC conducted by Jensen and Reynolds (1982) yielded three primary factors (verbal, spatial, memory) and only one second order factor, which is the *g*-factor. Another study of the Kaufman Assessment Battery for Children (K-ABC) with the 12 WISC subtests yielded the very same factor structure (Naglieri & Jensen, 1987). In another analysis by Carroll (1993), 400 data sets of cognitive ability test were analyzed and lead to his three-stratum model: Stratum I at specific level (for example, reading decoding, visual memory, sound-frequency discrimination); Stratum II at broad level and included eight factors: fluid intelligence, crystallized intelligence, general memory and learning, broad visual perception, broad auditory perception, broad retrieval ability, broad cognitive speediness, and processing speed; and Stratum III at the top level, a common factor very similar to the general intellectual ability *g*.

Different intelligence tests measure the g-factor of intelligence to varied degrees (Jensen, 1998), in other words, their g-loadings are different. The broad domains such as science are highly g-loaded; domains such as leadership are moderately g-dependent, while domains such as good citizenship have even lower g-loadings (Gottfredson, 2004).

Although g-loadings vary, g-factor had been shown to be stable across different collections of mental tests, even when tests bear little superficial resemblance (Jensen, 1986). Thorndike (1987) conducted a study with a pool of 65 highly diverse tests used in the armed services. He first randomly selected 48 tests to form 6 non-overlapping batteries of 8 tests each; then used the left-over 17 tests each as a ‘target test’, including them singly into the 6 batteries as a 9th test. This process yielded 6 g-loadings for each of the 17 tests. High correlations in the results showed that for each test, the g-loading stayed relatively invariant regardless of what batteries it is in; and the composite g-loading was close to its hypothetical value, even with the small numbers of tests used in the study.

All IQ tests were found to be highly g-loaded, and the practical predictive validity of psychometric tests mainly dependent on their g-loading. Correlational studies has linked g-factor with measures of biological constructs such as degree of inbreeding depression (Agrawal, Sinha, & Jensen, 1984).

Theoretical Critiques of IQ Tests

It is worth noting, that correlational evidence does not warrant *g* as a synonym for intelligence, and *g* did stem from a psychometric construct. As Spearman (1927, p.75) pointed out, “[t]his general factor g, like all measurements anywhere, is primarily not any concrete thing but only a value or magnitude. … Eventually, we may or may not find reason to conclude that g measures something that can appropriately be called “intelligence.” Such a conclusion, however, would still never be a definition of g, but only a “statement about it” (Spearman, 1927, p. 75-76).

Decades later many researchers still believe in the power of *g*, and considered it the main component of the intelligence construct (Gottfredson, 2004); but there are challenges as to whether intelligence can be equated to g-factor, or high g-loading tasks. One such challenge came from Howard Gardner, who was not satisfied with the definition of intelligence which is highly g-loaded. He believed that the g-loaded concept of intelligence and the resulting uniform school system was a concept too narrow, and formulated a list of seven intelligences. The list initially had seven types of intelligence: linguistic intelligence, logical-mathematical intelligence, musical intelligence, bodily-kinesthetic intelligence, spatial intelligence, interpersonal intelligence, intrapersonal intelligence; an eighth naturalist intelligence was later added to the list (Gardner, 1983). The first two types have been typically valued in traditional schools systems. Some scholars considered the first four intelligences to be in correspondence to some of Carroll’s Stratum II abilities (Gottfredson, 2004).

Gardner’s criteria for ‘intelligence’ includes as a prerequisite, the ability to resolve genuine problems or difficulties within certain cultural settings. But making judgments about it was, however, “reminiscent more of an artistic judgment than of a scientific assessment” (Gardner 1983, p. 63). Gardner claimed that these intelligences rarely function independently, but are used at the same time complementing each other as one navigates the environment, solves problems and develop his/her skills.

The dispute did not seem to be about whether *g* is a valid construct---*g* has proven to have strong link to IQ tests. The challenge on g-loaded definition of intelligence, boiled down to whether the IQ tests covered the full spectrum of human intelligence. If IQ tests could not fully measure human intelligence, then g-factor is likely to be a partial representation as well.

Correlations found between IQ and academic achievement have led scholars to believe that IQ test provide the best measures available of quality of human intelligence. High IQ was believed to be a good indicator of intellectual giftedness and high achievement (e.g. Erwin & Worrell, 2012; Naglieri & Bornstein, 2003; Neisser et al., 1996). However, predictive validity didn’t guarantee a causal relationship--this applies to the relationship of intelligence and achievement measures as well. Sternberg (1996) argued that what intelligence tests measures is itself one kind of achievement, developed under external influences; and Ceci and Williams (1997) showed that schooling has a direct impact on IQ scores. The overlapping component has raised challenges to the traditional concept of a *natural-born* intelligence, and its definition and measurement that bases solely on IQ tests.

As Binet (1916, p. 44) noted in his pivotal work, “some psychologists affirm that intelligence can be measured; others declare that it is impossible to measure intelligence. But there are still others, better informed, who ignore these theoretical discussions and apply themselves to the actual solving of the problem”. The contest between theory and practice will likely to continue, but the tension will positively help us gain deeper and more comprehensive understanding of what constitute human intelligence.

Developmental Critiques of IQ Tests

The traditional view towards intelligence not only considered it to be one faculty, but also one static trait. For example, Terman (1912) believed that a child’s IQ remains relatively constant at different ages. These views were challenged by researchers who held a more developmental view towards intelligence. Some scholars came to view intelligence and giftedness as dynamic, contextual and emergent (Dai & Renzulli, 2008). Simonton (1999) proposed that giftedness is relative to a specific domain that offers a specific set of opportunities and challenges to an interested person. Whether gifted behavior will emerge depends on the nature of the domain and the interaction of the domain and person. Thus giftedness is not a trait bound to emerge, but a critical state fostered in important aspects of development (Ziegler, 2005).

Cultural Factors about Intelligence

If external or environmental factors are considered as a factor for intelligence development, culture has to be one salient component. Anastasi wrote in 1992, “Intelligence is not a single, unitary ability, but rather a composite of several functions. The term denotes that combination of abilities required for survival and advancement within a particular culture" (p. 613). In his attempts to establish a construct of intelligence to serve a broader purpose, Sternberg (1985) took into account both personal and cultural notion of success, and defined intelligence as one's ability to achieve success in life in terms of one's personal standards, within one's socio-cultural context. His theory of success intelligence states that a balancing of abilities is achieved in order to adapt to, shape, and select environments; and the triarchic theory postulated a separation of analytical intelligence, creative intelligence, and practical intelligence. In this theory, g-factor was considered one of the three independent intelligences, although there has been other independent study that showed high correlations between the three intelligences (Brody, 2003).

Models aside, society is inevitably influenced and adapt to what was practiced and valued around us. Intelligence may be conceived in different ways in different cultures (Sternberg & Kaufman, 1998; Yang & Sternberg, 1997). Different ethnic groups in the same physical area can have different conceptions of what intelligence means (Okagaki & Sternberg, 1994). Contextually important skills are so critical in certain environments, that children may develop these skills in place of academic ones; their substantial skills could be overlooked in academic tests (Sternberg & Grigorenko, 2004) . And even when different cultural groups were found to have similar underlying structure of the concept of intelligence, there is a difference in their perception towards the ranked difficulties of certain skills (Yang & Sternberg, 1997). Sternberg (2007) thus argued that there is no one overall conception of intelligence.

Relating to Personality

Researchers believed that for achievement that is long term or requires everyday performance, personality variables should be considered along with IQ scores as predictors (Chamorro-Premuzic, Moutafi, & Furnham, 2005). Giftedness is one of such measure. Taking personality traits in account is likely to provide us more information in the process of identifying gifted or potentially gifted students.

The association of personality traits and intelligence has long been of interest to researchers and a plethora of studies have been conducted (e.g., Ackerman, 2009; Chamorro-Premuzic, Moutafi, & Furnham, 2005; Saklofske & Zeidner, 1995). In the long line of development in this field, these extensive reviews, along with others, helped to provide some basic ideas of the progression in understanding the relationship of personality traits and intelligence. A comprehensive review of these studies is beyond the scope of this chapter.

Personality models. In contrast to a widely accepted general factor (g-factor) in intelligence, in the realm of personality traits, no single general factor was found. Rather, a hierarchically organized structure with several broad personality traits was generally agreed upon by personality researchers. Specific personality traits that would vary in a similar fashion are housed under an umbrella of each broader personality traits. The Big Five, or the Five Factor Model, is one model that many researches converged upon. The five factors include: Neuroticism, Conscientiousness, Extraversion, Agreeableness, and Openness, and each of these five domains comprises of a large number of lower-level traits known as facets. For example, nurturing, caring, emotional supportive are lower level traits under agreeableness, while hostility, jealousy, indifference, are under the same domain at the opposite end. There is no consensus as to the number of constituent traits for each domain.

Differences and relations between intelligence and personality. Intelligence and personality has traditionally been viewed as two separate entities. This dichotomy was naturally derived from many differences in these constructs, such as in their trait and measurement differences (DeYoung, 2011; Most & Zeidner, 1995), detailed in the follows.

- Trait Differences include:
    1. Intelligence has been viewed as unidirectional, with the optimal parameter set at the higher end of the spectrum, while on the other hand, personality traits, on the other hand, could be considered bipolar (or bidirectional). Parameters at both end of personality extremes are

considered non-optimal while an optimal value could be hard to define.

2. Intelligence is usually associated with cognitive abilities, and personality is associated with non-cognitive features.

- Measurement related differences include:
    1. In intelligence tests where individuals are asked to complete a task or tasks at a set time and environment, they understand that they are being evaluated and are expected to output at their highest level. In contrast to 'maximum performance' for intelligence test, personality assessment usually concerns with 'typical behavior'. Some might also argue that there is more voluntary control in modifying one's personality states than intelligence states.
    2. A difference in assessment methods exist between the two: intelligence is usually tapped by standardized ability tests while personality is probed by behavior-ratings, self-reports, questionnaires, etc.
    3. Personality measures tend to have lower reliability and validity than intelligence measures. They are more likely to have measurement errors from multiple sources, and their interpretation is more ambiguous than intelligence measures.

Assessments and theories in personality seemed to be less unified than intelligence testing, possibly because no personality traits are considered as dominant in the personality domain as academic intelligence in the intelligence domain. Some critique to the personality literature included that it is “rife with isolated personality measures of varying levels of breadth, often with no linkage to any personality theory” (Ackerman, 1997, p. 222).

On the other hand, efforts to integrate personality and intelligence into a coherent framework were made. For example, Furnham (1995) proposed a model that related personality, intelligence and academic achievement in the same model. Personality and intelligence were both viewed as predictors to academic achievement, and also to cognitive/learning styles (CLS), which itself serves as a moderator. Teaching and assessment methods are considered to be independently related to CLS. On the more integrative side, some researchers went as far as to interpret intelligence as part of personality. DeYoung (2011) challenged many dichotomies between intelligence and personality and contended that intelligence can possibly be fit in the Big five scale. Lohman and Rocklin (1995) also pointed out that the distinct boundary set between traditional intelligence and personality or affectivity was set up for the convenience of research, and do not fully reflect the process of a student’s growth.

Motivation and Creativity in Giftedness

Two particular traits are frequently addressed with giftedness: motivation and creativity. Terman(1916) addressed *emotion and volition* as limitations of the Stanford-Binet scale, but more contemporary researchers (e. g. Dai & Sternberg, 2004; Shavinina & Seeratan, 2004) argued that emotions and motivations cannot be cleanly separated with cognitive capacity and processes. As a matter of fact, affect and motivation were considered more important than cognitive abilities for success in certain domains (Winner, 1996).

Similarly, creativity was another challenge often posted against IQ testing. The traditional concept of intelligence was measured mainly by reading, mathematics and similar tasks, and did not tell us much about a person's creative nature (e.g., Guilford, 1950). Creativity was often described as a personality, and strongly associated with the Openness part of the Big Five model; at the same time however, it has also been termed in intelligence, again demonstrating the blurred line between intelligence and personality. In Sternberg's triachic theory (1985), creativity was termed as creative intelligence, along with analytical intelligence and practical intelligence. No matter what it is classified as, the importance of creativity in models of giftedness is evident.

Renzulli (1986) pictured a three-ring model of giftedness, where he identified three important factors: above average ability, creativity, and task commitment. Even though separate, these three factors connect to contribute to a person's giftedness. A distinction was made between general abilities and specific abilities. The developers of HBGSI broadly borrowed from Renzulli's basic definition of giftedness and expanded it to the Hispanic bilingual gifted student as "one who has above average intelligence (IQ), task commitment, and creativity that is situated within socio-cultural-linguistic characteristics" (Lara-Alecio & Irby, 1993, p. 6).

A Legal Definition

Even though there has been no consensus on intelligence and giftedness, some definitions were followed more frequently in practice. National Association of Gifted

Child (NAGC, 2008) noted as one of its “non-negotiable practices” that the choice of assessment tools must match the definition of giftedness that has been determined by the state, district, or school. In accordance to the Texas State Plan for the Education of Gifted/Talented students (2009, §29.121, p.11), the term *gifted and talented students* means:

A child or youth who performs at or shows the potential for performing at a remarkably high level of accomplishment when compared to others of the same age, experience, or environment and who:

1. Exhibits high performance capability in an intellectual, creative, or artistic area;
2. Possesses an unusual capacity for leadership; or
3. Excels in a specific academic field.

In this definition, evidence of influences from multiple theories about intelligence and personality can be traced. Still, some aspects were not stressed. For example, it seemed to have placed little measure on developmental potential and was mainly performance based.

Under-representation of Hispanic Bilingual Students

Hispanic bilingual students are among several of the minority groups who were underrepresented in G/T programs. The situation called for comprehensive examination at this complex issue rather than expectation of a single silver-bullet solution (Callahan, 2005).

Demographics

Hispanic population has been the fastest-growing population in United States (U.S. Census, 2010). Percentage wise, from 1999 to 2009, the proportion of public school enrollment accounted for by Hispanic students increased from 15.6% to 22.1% (6.5% points increase) nationwide (U.S. Census, 2010). Over the same period, the number of students receiving bilingual or English as a second language (ESL) instructional services grew by 49.2%, and the number of students identified as limited English proficient (LEP) grew by 39.4%.

In the school year of 2012-2013, 2,606,126 Hispanic students were enrolled in public schools in Texas, accounting for 51.3% of total enrollment. This is the largest percentage ethnic-wise, followed by White (30.0%), African American (12.7%), and Asian students (3.6%). In the same year, a total of 16.2% (809,854 out of 4,998,579) of the students enrolled were receiving instruction services in bilingual or English as a second language. Out of these students, 90.5% of them are Hispanics.

In comparison, in gifted and talented programs, Hispanic students accounted for only 40.6% of the overall enrollment, in the academic year 2012-2013 (TEA, 2014), less than its overall student representation percentage. Note that this percentage was an increase from what was reported about a decade before: 28.2% of all GT enrollment in the year 2001 to 2002 (TEA, 2001, p. 19). Data that reported students who were both bilingual Hispanic students and were in gifted programs was difficult to locate, suggesting possible oversight of these students as a group. In the words of TEA (2008, para. 2), “discrepancies still exist between the percentage of underrepresented populations in the total student population versus the percentage of underrepresented populations identified for G/T services."

Identification and Assessment for Hispanic Bilingual Students

Researchers have called for instruments that are valid and reliable, and multiple of such sources should be referenced as evidence for decision making (Callahan, 2005; NRC, 2002; NAGC, 2008; Worrell, 2009). Identification and assessment of potentially gifted LCD studentsare difficult because complex cultural and linguistic background differences exist between these students and the native English-speaking children from middle-class families (Cohen 1998). Language creates barriers for non-native speakers. In some cases, ELLs who do not speak English at home, were tested in English at school (e.g. Harris, Plucker, Rapp, & Martinez, 2009). In such practices, as the test is not conducted in the students' dominant language, the test is also assessing students' English language proficiency besides the target constructs (Standards for Educational and Psychological Testing, 1999). In fact, any test using a language is in some way measuring language skills, but they put students who were still learning the language at a greater disadvantage. The fact that Hispanic Bilingual students were tested by general intellectual abilities tests in English, was noted as one cause of underrepresentation of these students in GT programs (Ouyang & Conoley, 2007). Harris, Rapp, Martinez, and Plucker (2007) contended that Hispanic bilingual students are unlikely to be identified if they are assessed with standardized tests that are only given in English.

Nonverbal tests provide students opportunities to demonstrate intelligence without influence confounded by language, vocabulary, and academic exposure (Ford &

Graham, 2003). Such tests include the Raven's Standard Progressive Matrices (RSPM, Raven, Raven & Court, 2000), Naglieri Non-verbal Ability Test (NNAT, Naglieri, 1997), as well as the Wechsler Nonverbal Scale of Ability (Wechsler & Naglieri, 2006). In a study by Saccuzzo, Johnson & Guertin (1994), substantively more Hispanic and Black students were identified as gifted when Raven's tests were given than when traditional tests (WISC-R) were used. Naglieri and Ford (2003) reported that the difference in mean score and percentage of students obtaining high standard scores were small, when students were tested by NNAT.

Extending the Naglieri Non-verbal Ability test, Naglieri and Das (2005) followed the Planning, Attention, Simultaneous, and Successive (PASS model (Das, Naglieri & Kirby, 1994; Suzuki and Valecia, 1997), and created the Cognitive Assessment System (CAS). Standard scores are provided for each section as well as the full scale. The traditional progressive matrix tests similar to NNAT (Naglieri, 1997) were included in the 'Simultaneous' part. Naglieri, Rojahn, and Matto (2007) measured Hispanic and non-Hispanic groups using CAS, and found small differences between ethnic groups (negligible when parental education level is statistically controlled). Their results supported previous expectations (Fagan, 2000; Suzuki & Valencia, 1997) that difference between Hispanic and non-Hispanic children would be relatively small, when the ability test focus on basic cognitive processes. In this study, contrary to the general belief that a more inclusive measurement achieve racial equality at the cost of psychometric rigor (e.g., Gottfredson, 2004), the reduction in mean score differences between the ethnic group classifications did not appear to have come at the expense of reduced validity.

Teacher Preparation

As suggested by the above study, redefining the concept of intelligence may reduce the differences between majority and minority groups, and provide a comprehensive way to conceptualize and measure ability. This is true outside the assessment society and practical for school environments too.

The overly narrow understanding of giftedness might be one obstacle preventing culturally linguistic diverse students from entering gifted programs (Ouyang & Conoley, 2007). As GT identification begins in most instances in the classroom (Castellano, 2011), classroom teacher is a big factor in nominating students for GT programs (Irby, Lara-Alecio, & Rodriguez; 2003). However, few teachers are exposed to multicultural educational experiences, nor practical training in multicultural curriculum and instruction in urban settings (Banks & Banks, 1995). Minority teachers are also found to have little exposure to gifted education in their preparation (Ford, 1999), let alone the various concepts of intelligence discussed in earlier sections. Teachers ability to make fair and equitable referrals are impeded by their lack of prepared knowledge and inadequate sensitivity to the characteristics of minority gifted students, by their insufficient understanding of the social and emotional needs of gifted students, especially the underachievement among them (Ford & Graham, 2003). For example, in a study by Plata, Masten, and Trusty (1999), Hispanic students nominated for GT programs resembled the characteristics of the nominated Anglo students.

Many characteristics are unique for different groups, and Hispanic bilingual students have characteristics of their own. As pointed out by Ford (2012), the term ‘minority students’ are too often discussed and treated as one homogenous group, and researchers (e.g., Oghu, 1992; Oghu & Davis, 1993) urged educators to not discount or minimize fundamental differences between various racial and ethnic groups. Characteristics manifested in Hispanic bilingual students’ learning, communication, and behavioral styles, are shaped by unique cultural influences. But they can be misunderstood by teachers, or even viewed as deficits (Ford, 2012). For examples, Puerto Rican children learned to resort to their family for advice, rather than act independently (Perrone and Aleman, 1983); Mexican children learned to respect elders and authority but may be vulnerable if the school system emphasized values such as individual competition, initiative and self-direction (Cohen, 1988); Hispanic students found to benot normally assertive and non-authoritarian in comparison to Anglo students, could be at risk of being left out from GT programs (Plata, Masten, & Trusty, 1999). Therefore, it is important that we can identify the characteristics of Hispanic bilingual gifted children, as well as promote awareness in teachers and school systems of these identifiable traits.

Item Response Theory (IRT) and Its Applications

IRT has been recognized as an important framework in instrument construction and validation, and provide multiple advantages over CCT (Embretson & Raise, 2000). Within the IRT framework, we can examine psychometric properties in a different perspective from CCT: validity (dimensions), reliability (item and test information), and item selection based on item properties. In particular, IRT has shown to be useful in refining and shortening psychological instruments (raise & Waller, 2009).

Basic Concepts and Unidimensional IRT

The following paragraphs briefly describe basic principles of IRT and the unidimensional model. In a nutshell, the core function in an IRT theory depicts the probability of a particular item response given a certain trait level. This function usually plots an S-shaped form, and can be defined in various ways, such as the logistic function, the normal ogive function, or even a step function.

Dichotomous IRT model. Dichotomous responses commonly exist in educational research: they are just items where answers can be scored as either 0 or 1. The item itself can be a question with multiple types of answers, but the answers are eventually coded into two types, as in the example of how multiple choice answers are sometimes coded as correct and incorrect in the scoring process. For dichotomous items, a mathematical function is chosen to describe the probability of a person endorsing a particular item, meaning that the person's response is 1 instead of 0. The item itself can be a question with multiple types of answers, but the answers are eventually coded as either 1 or 0.

Take the logistic function as an example. According to Embretson and Reise (2000), the basic one-parameter logistic (1PL) model uses the distance between person ability and item difficulty as the parameter. This model is also known as the Rasch model. The mathematical formula is as follows:

$$P\left(X_{ij} \ = 1 \middle| \theta_i\right) = \frac{e^{\theta_i - b_j}}{1 + e^{\theta_i - b_j}}$$

$P(X_{ij} = 1|\theta_i)$ is the probability of the *i*th person endorsing the *j*th item, while θ_i denotes level of the latent construct of *i*th person, and b_j denotes the level of difficulty of the *j*th item. b_j is also referred to as the difficulty parameter, describing what is sometimes referred to as the location of the item. In IRT, the function is also known as the item response curve (ICC).

On top of the 1PL model, we can add another item parameter *a* to describe how well an item discriminates among different ability levels. The two-parameter logistic (2PL) model is mathematically defined as follows:

$$P(X_{ij} = 1|\theta_i) = \frac{e^{a_j(\theta_i - b_j)}}{1 + e^{a_j(\theta_i - b_j)}}$$

The denotations are similar to those in the 1PL model formula. The additional parameter a_j is known as the discrimination parameter in IRT literature, indicating the slope of the ICC. A higher value in discrimination parameter indicates bigger change in probability with the same amount of change in ability, thus is usually preferred over lower values as it indicates stronger relation between the item and the latent construct. Further parameters can be added to the model, forming 3PL or even 4PL models, and they are all nested models.

Polytomous IRT model. Polytomous items can be modeled by IRT as well. When item responses consist of ordered categories, the same rational as in binary models applies to polytomous IRT models. According to Ostini and Nering (2006), multiple IRT models are available, such as the Partial Credit Model (PCM, Masters, 1982), the General Partial Credit Model (PCM, Masters, 1992), and the Graded Response Model (GRM, Samejima, 1969).

Essentially in polytomous IRT models, the questions regarding probabilities of each ordered categories are decomposed into a set of questions with a binary answer concerning the relationship between adjacent categories. For example, in PCM, an item is broken down into several steps. Each step is the relative difficulty between adjacent items: how difficult is it to go from one category to the next---with difficulty described by probability modeled by the core function. Rating scale model can be viewed as a special case of PCM where all items have the same number of categories and all relative difficulties between adjacent categories are equal. In PCM, the transitional probability from one category to the next category is expressed by the same logistic function in Rasch model. The plot of this function is known as category response curve, and an item with *k* categories would have *k-1* category response curves. When we extend this 1PL model to a 2PL model, to describe the same transitional probabilities, we get the General Partial Credit Model (GPCM). In both these models, the difficulty parameter is the point of the latent variable scale where two adjacent category response curves intersect. Mathematically it is possible that step difficulties are not in the same order as in the category, which conceptually maps to cases where initial steps are harder than proceeding ones. In practice, this might happen if little differentiation exists between category responses in some personality or attitude measures.

While steps in PCM models can be independent, they are progressive in Samejima's Graded Response Model: endorsing one category does require all lower categories being endorsed. Unlike PCM which models conditional probabilities of each step given its previous step in PCM, in Samejima's GRM, it is the cumulative probability for each category and its higher categories that is modeled by the core function. In other words, the decomposed dichotomous question is: for each category (*k*), 1 means that the response is in this category (*k*) or higher (*k+1* and higher); and 0 means that it is below this category (*k*-1 and lower). The probability of endorsing a particular category (*k*), therefore, is the difference of the cumulative probabilities of the next higher category (*k+1*) and itself (*k*). The item difficulty parameters are known as threshold parameters, and are always in the same order as the categories themselves. At the two extreme ends, thresholds is the point where the probability is .50 for the lowest and highest category, while in between, it is determined by the adjacent threshold and the peak of the category curve.

Logistic and normal ogive functions. Even though logistic function was used as an example in formulas presented earlier, both the logistic function and the normal ogive function can be used as the core function (a.k.a. link function). Let's take 2-parameter Graded Response Model as an example.

The logistic function is given by:

$$P\left(X_{ij} > k \middle| \theta_i\right) = \frac{e^{a_j(\theta_i - b_{jk})}}{1 + e^{a_j(\theta_i - b_{jk})}}$$

And thus for each category:

$$P(X_{ij} \ = k|\theta_i) = P(X_{ij} \ > k|\theta_i) - P(X_{ij} \ > k+1|\theta_i)$$

$$= \frac{e^{a_j(\theta_i - b_{jk})}}{1 + e^{a_j(\theta_i - b_{jk})}} - \frac{e^{a_j(\theta_i - b_{j(k+1)})}}{1 + e^{a_j(\theta_i - b_{j(k+1)})}} = \frac{e^{a_j(\theta_i - b_{jk})} - e^{a_j(\theta_i - b_{j(k+1)})}}{(1 + e^{a_j(\theta_i - b_{jk})})(1 + e^{a_j(\theta_i - b_{j(k+1)})})}$$

The normal ogive form of the graded response model is given by:

$$P(X_{ij} \ = k|\theta_i) = \frac{1}{\sqrt{2\pi}} \int_{a_j(\theta_i - b_{j(k+1)})}^{a_j(\theta_i - b_{jk})} e^{\frac{-t^2}{2}} \mathrm{d}t$$

Logistic function and normal ogive functions can be very similar, as shown in the following Figure 1 (Wolfram Demonstration Project, 2014), in the two almost identical curves. In this figure, the red curve represents the ogive function, and the blue curve represents the logistic function, with all parameters of the same value, except for the discrimination parameters, which is scaled up by 1.701. These two functions differ by less than 0.01 across the continuum.

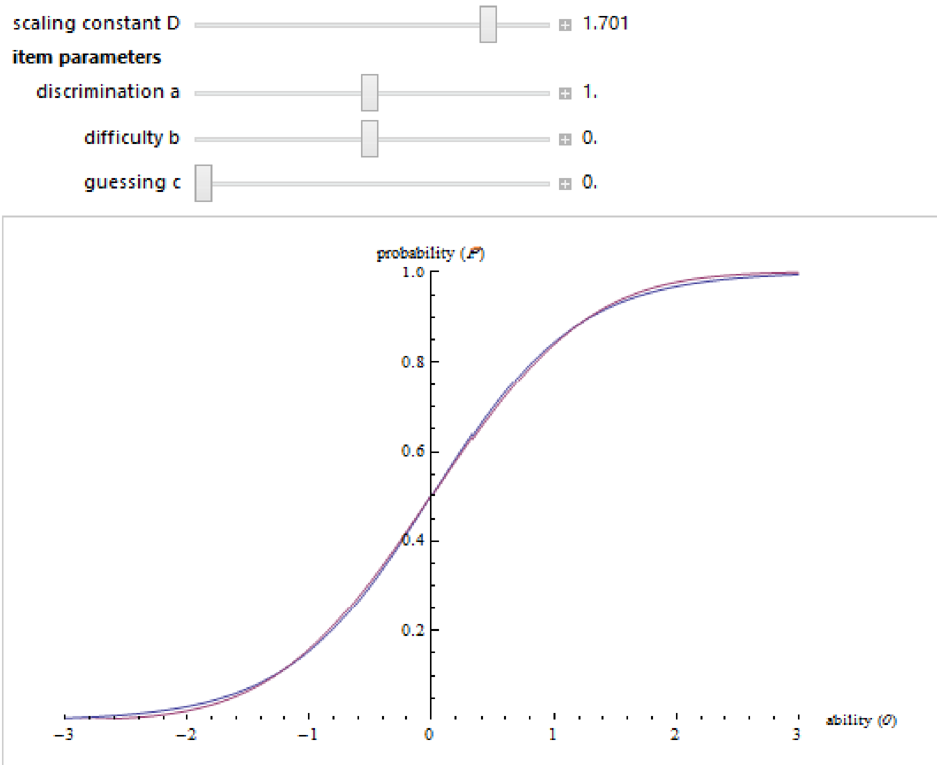

Figure 1. Logistic and normal ogive functions (Wolfram, 2014)

Consequently, results modeled by the logistic function and normal ogive function are very similar, when a scaling constant is applied to the discrimination parameter in logistic models. Conventionally a scaling factor D =1.7 is used in the logistic formula, and the formula are sometimes rewritten as

$$P(X_{ij} = k|\theta_i) = \frac{e^{Da_j(\theta_i - b_{jk})} - e^{Da_j(\theta_i - b_{j(k+1)})}}{\left(1 + e^{Da_j(\theta_i - b_{jk})}\right)\left(1 + e^{Da_j(\theta_i - b_{j(k+1)})}\right)}$$

to have comparable results with normal ogive model. The choices between models are closely related with computational methods, rather than conceptual differences.

Test statistics. A collection of items forms the instrument of measurement, commonly referred to as a test in IRT. There are two important test traits and they are depicted by two distinct curves: the test characteristic curve (TCC), which is the expected value of the summed score plotted against θ; and the information function curve (IFC), which is the sum of information provided by all items.

Test information is a crucial concept in IRT, and is used to describe the precision of measurement, analogous to the concept of reliability in classical test theory (CTT). Because items in IRT models measure trait level with varied quality across the continuum, item information is usually represented by a function instead of a constant. Item information function is positively related to item discrimination parameter (function-wise) and is inversely related to the variance of the trait level estimate θ. Thus standard error is the reciprocal of the square root of information, and varies as information varies across the continuum. Test information is the sum of all item information because of the local independence assumption. Information function for GRM was derived by Samejima (1969) and plot to be a multimodal curve.

Multidimensional IRT

According to Reckase (2009), multidimensional IRT can be differentiated into two types. The first type is typically called compensatory models as probability of responses are based on a linear function of combination of θ values, combined from a vector of θ coordinates representing the multiple dimension. Another type is commonly known as the noncompensatory model, where a task is broken into several parts and each part is modeled unidimensionally. The probability of correct response from the item is the product of the probability of each part. Because product is used instead of addition in the second model, nonlinearity exist in this model, and θ value resulted from one dimension cannot replace value from another dimension and, is thus ‘noncompensatory’. In the following sections, only compensatory model is explained.

Extension from GRM. One way to extend UIRT to MIRT, is to represent multiple dimensions with vectors, instead of scalers in UIRT. Specifically, in a MIRT model with m dimensions, the product of $a_j\theta_i$ can be replaced with $\mathbf{a}_j\boldsymbol{\theta}_i$, where $\boldsymbol{\theta}_i$ is a $1 * \mathrm{m}$ vector with each member representing person’s location in regard to the m dimensions repectively, and $\mathbf{a}_j$ is a $1 * \mathrm{m}$ vector correspondingly composed of discrimination parameter in regard to the m dimensions.

For a simpler example in dichotomous items, 2PL model can be written as

$$P\left(X_{ij} = 1\middle|\boldsymbol{\theta}_i\right) = \frac{e^{\boldsymbol{a}_j\boldsymbol{\theta}_i+d_j}}{1 + e^{\boldsymbol{a}_j\boldsymbol{\theta}_i+d_j}}$$

The intercept d_j is a scalar and is similar to the intercept in the UIRT model, which equals to -$a_j b_j$. If $\boldsymbol{\theta}_i$ is 0, the probability of person *i* scoring correctly is 0.5. If all other elements in $\boldsymbol{\theta}_i$ is 0 except the *l*th element, the probability of correct response for a test item is $-d_j/a_l$.

Note that because of the existence of compensatory dimensions, intercept on each axis does not fully capture the location of the item. The location, or difficulty parameter is best described by the distance between the origin and the line, which is expressed as

$$b_j = \frac{-d_j}{\sqrt{\boldsymbol{a}_j \boldsymbol{a}_j'}}$$

Similarly, because the model extends from unidimensionality to multidimensionality, item and test characteristics represented by curves in UIRT are represeted by surfaces in MIRT. Visual representations of curves in 2D space are represented by multidimensional surfaces and surface countour plots in 2D spaces.

In a similar fashion, for polytomous items, the graded response normal ogive model can be expressed as

$$P(X_{ij} = k | \boldsymbol{\theta}_i) = \frac{1}{2\pi} \int_{\boldsymbol{a}_j' \boldsymbol{\theta}_i + d_{j(k+1)}}^{\boldsymbol{a}_j' \boldsymbol{\theta}_i + d_{jk}} e^{\frac{-e^2}{2}} \mathrm{d}t$$

where $\boldsymbol{a_j}$ is the vector ofdiscrimination parameters of item j, and d_{jk} is the parameter related with the *k*th category of the item similar to the case as in logistic model.

MIRT and Item Factor Analysis (IFA). The compensatory model is closely related with factor analysis, even though Item Response Theory (IRT) and Factor Analysis (FA) have its distinct origins and traditions. Early work (e.g. Takane & de Leeuw, 1987, Muthén, 1984, Bock, 1988, Bartholomew, 1983, 1985) has established mathematical equivalence of the two-parameter normal ogive model in item response theory and factor analysis for both dichotomous and polytomous variables with marginal likelihood. For ordered categorical data, the logistic model are considered a close alternative to categorical factor analysis, if it provides a good approximation to the ogive model (de Leeuw, 1983, Takane & Leeuw, 1987). The dichotomy between IRT and

Factor Analysis have been attenuating and it is much more common in recent literature that IRT are also considered as categorical l item factor analysis (Baker & Kim, 2004; Fox, 2010, du Toit, 2003, Muthén & Muthén, 2012). Despite the virtually identical process, multidimensional IRT (MIRT) and Factor Analysis differ in the following three ways (Reckase, 2009) :

1. Goal: FA is thought as a data reduction process while MIRT models the interaction between persons and humans.

2. Focus: MIRT focus on characteristics of input variable (items), while FA generally ignores them.

3. Methods: MIRT analyses work towards finding solutions in a common coordinate system in the latent space, across samples and tests, and look for methods that extend the unidimensional IRT to MIRT, in regard to test linking, in equating and calibration; development of item bank; and computer adaptive testing. In comparison, FA now emphasis on confirmatory methods and structural equation modeling.

IRT for Instrument Construction

IRT has found application in personality tests, clinical tests, intelligence tests, and other forms of measurements, and in various fields such as psychology, public health, organizational management, as well as intelligence tests in education. Several major intelligence tests have applied IRT, such as the Differential Ability Scales, the Woodcock-Johnson Psycho-Educational Battery, the Stanford-Binet test (Embretson & Reise, 2000). Many applications of IRTs can be found for smaller scale instruments too, but there is a lack of literature of IRT used for instruments that identify Hispanic Bilingual gifted and talented students.

IRT can be utilized in multiple aspects of measurement evaluation and construction, including:

1. Instrument construction and adaptation : items can be selected by examining their validity (proper dimensionality), reliability (information), adequate variance, and adequate item difficulty and discrimination

2. Instrument evaluation: IRT can be used to evaluate measurement bias for certain groups, as we can check if the same model apply across groups. In IRT model, measurement is biased if differential item functioning (DIF) exists for the groups being compared. Software packages usually provide procedures to evaluate DIF.

3. Test scaling and Equating.

4. Computer Adaptive Testing.

For the purpose of this study, I elaborate on one particular application that IRT has shown to be particularly useful and powerful: refining and shortening psychological instruments (Reise & Waller, 2009). Compared with classical test theory, IRT identify items with most information and discrimination properties relative to other items, and provide information for decisions in shortening scales with minimal loss of information (Reise & Henson, 2003).

Ebesutani et. al (2012) conducted a study of an instrument named Loneliness Questionnaire (LQ; Asher, Hymel, & Renshaw, 1984), with a large sample of 6,784 children and 4,941 adolescents. They selected items based on multiple criteria, including items associated with high reliability, high IRT discrimination (slope) parameters, high item or test information, and location properties demonstrating adequate coverage and discrimination across the various levels of loneliness. Their results showed that non-reverse-worded items provide more information than the reverse-worded items, and therefore provided evidence to eliminate the non-reverse-worded item from the instruments. The authors proposed a reduction of items in the instruments, as well as reduction of categories for the item responses.

Bann et al (2012) used responses from 5399 survey participants for the purpose of developing and evaluating a shortened version of the Public Health Surveillance Well-Being Scale (PHSWB), an instrument that measure mental, physical and social components of well-being. Three analyses were conducted to exam the psychometric properties of the original 34 items: a bi-factor analysis (Schmid & Leiman, 1957), IRT analysis, and DIF analysis. The authors chose 10 best items based on the following criteria: (a) no exhibition of ceiling or floor effects; (b) varied item responses; (c) missing data rate is not high; (d) high factor loadings (>=0.40) and IRT slopes (>=1) ; (e) items should not demonstrate slope related DIF. The shortened version of PHSWB showed high correlation with the original instrument, as well as good internal consistency. These psychometric properties provided good evidence for minimal loss of information with the cut in item numbers, and allowed future wellbeing assessments be conducted in shorter forms, such as in national surveys.

In this study, I applied IRT to examine the psychometric properties of Hispanic Bilingual Gifted and Talented Screening Instrument (HBGSI) at item level, and explored the possibility of a shortened version.

CHAPTER III

METHODS

Introduction

As described in previous chapters, the increasing population of Hispanic students called for better instruments in identifying Hispanic Gifted and Talented (G/T) students. Due to complexity of the concept of intelligence and giftedness itself, no single tool can serve as a panacea for G/T identification, and a comprehensive set of instruments are recommended to prevent potentially G/T Hispanic children from being denied of appropriate education. In alignment with the purpose of HBGSI, I focused on the first phase of the identification process, i.e., the teacher referral process, where students are screened and then referred for further GT testing.

The purpose of this study was to further validate HBGSI within the framework of IRT, exploring the factor structure and dimensionality of the instrument at the item level. I further tested the possibility of constructing an abbreviated version of HBGSI with fewer items for ease of administration, which would potentially lower the demand on the teacher's time, enhance accessibility and facilitate increased usage of the instrument.

Context

Participants of this study were part of a randomized control trial, known as English Language and Literacy Acquisition (ELLA). ELLA project was sponsored by Grant R305P030032 from Institute of Educational Sciences (IES), the research arm of the U.S. Department of Education. This longitudinal project followed predominately Spanish-speaking English language learners (ELLs) from kindergarten to grade 3 in an urban school district, in an effort to deliver quality ESL instructions in both Structured English Immersion (SEI) and Transitional Bilingual Education (TBE) programs to promote ELLs' English language and literacy, and also to compare effects when instructions are delivered in high quality, and in a fashion reflecting more typical practices.

In transitional bilingual education models, the focus shifts from oral language development in English in kindergarten, to content instructions in Science and Social Studies in grade 3. The typical practice of TBE in the school district begins with an 80% (Spanish) / 20% (English) model in Kindergarten and gradually shifts to a 50/50 model in grade 3. There is a 45-minute ESL component with no support from ELLA research team. The enhanced practice of TBE in the school district is a program that begins with a 70% (Spanish)/ 30% (English) model in kindergarten and progressively moves to a 40/60 model in grade 3, increasing portion of English at every grade. In kindergarten, the 75 minute ESL component consisted of daily tutorials in Intensive English (Ventriglia & González, 2000) program, Story Telling for English Language and Literacy Acquisition ([STELLA], Irby, Lara-Alecio, Quiros, Mathes & Rodriguez, 2004), and teacher-conducted Daily Oral Language using Question of the Day (Lakeshore, 1997). Starting from grade 1, ESL intervention time increase to 90 minutes and includes Santillana Intensive English / Interactive Writing (Ventriglia & Gonzalez, 1999), STELLA, and Science-based Oral Language Development. Enhanced transitional bilingual education program in effect achieves a one-way dual language program because of the following characteristics: (a) subject matter is taught in the first and/or second language; (b) literacy is developed in the first and second language; and (c) comprehensible input is provided in English and the second language (Irby, Tong, Lara-Alecio, Mathes, & Rodriguez, 2004). In third grade, Content Reading Integrating Science for English Language and Literacy Acquisition (CRISELLA, Irby et al, 2008) were in place with strategies to help second language learners develop science academic language and expository reading skills.

The typical practice of SEI in the school teaches all subjects in English with rare clarifications in Spanish. There is a 45-minute ESL component with no support from research team of Project ELLA. In the enhanced practice of SEI, all subjects including content areas were taught using ESL strategies such as total physical response, visual aids, gestures, and other appropriate strategies. Language development strategies were included in the content subject. Spanish was used to clarify only when needed. Similar to the enhanced practice of TBE, the 75 minute ESL component in kindergarten consisted of daily tutorials in Intensive English (Ventriglia & González, 2000) program, STELLA, and teacher-conducted Daily Oral Language using Question of the Day (Lakeshore, 1997). Starting from grade 1, ESL intervention time increase to 90 minutes. For details of the intervention, please refer to the project report and publications by Irby, et al., (2010).

Twenty-four elementary schools in an urban district near Houston, Texas participated in this large experimental study. This district was one of the largest school districts in Texas. In the previous academic year when project ELLA started, 56.3% of the district's students were Hispanics, and about 41% of them (23.5% of the whole student body) were classified as Limited in English proficiency (LEP), a.k.a. English Language Learners (ELLs) (TEA, 2003). A majority of the students (74.2%) are economically disadvantaged (TEA, 2003). A total of 21.8% of students were enrolled in bilingual/ESL Education, in one of the three programs the district provided for ELLs at the time of study: structured English immersion, transitional bilingual program, and two-way immersion program (TEA, 2003).

The district earned 'RECOGNIZED' status in accountability rating by TEA in 2002-2003. This district was chosen for ELLA project because of long-standing reputation and experiences working with ELLs, its consistency in program philosophy and implementation, and access to SEI and TBE programs within the district (Tong, 2006).

Participants

All the students were identified by the state as ELLs at the beginning of kindergarten, and spoke Spanish as their native language. These students were placed in programs of either Structured English Immersion (SEI) or TBE (Traditional Bilingual Education). According to the final performance report of English Language and Literacy Acquisition Project (Lara-Alecio, Irby, Tong & Mathes, 2009), 822 students (experimental n = 464, control n = 358) students participated in the kindergarten school year; 768 students (experimental n = 394, control n = 374) participated in first grade, 517 students (experimental n = 261, control n = 256) were in the second grade and 390 students (experimental n = 188, control n = 202) stayed for the third grade.

My study included students from grade 1 to grade 3, whose item-level data were available at the time of study. Students in Kindergarten were not included because they were still at the very initial stage of development and exhibit different traits compared to when they are in the elementary grades. Table 1 shows the number of students in each grade.

Table 1

Number of Students in Each Grade

Grade	# of Students
1	383
2	396
3	338
Total	1117

Instrumentation

Content

HBGSI is an instrument designed to be used in the first stage of GT identification, for the teachers to determine if a student is potentially gifted and should be referred for further assessment for giftedness. This instrument was recommended for students from K to grade 4, with an adapted version for PK and K in progress. A full list of items is presented in Appendix A.

Giftedness was defined by the developers as “one who has above average intelligence (IQ), task commitment, and creativity that is situated within socio-cultural-linguistic characteristics” (Lara-Alecio & Irby, 1993). It was the developers’ belief that “to compare one ethnic group of gifted students to another ethnic group of gifted students is competitive, ethnically-biased research” (Irby, Lara-Alecio, & Rodriguez, 2003b, p.5), and was intended not to compare Hispanic students with mainstream population, or other minorities, but rather to identify the Hispanic, bilingual, gifted students given their socio-cultural-linguistic backgrounds.

A total of 77 items grouped into 11 clusters were included in the current version of HBGSI. Each cluster contains between 3 and 15 items. According to Irby, Lara-Alecio, and Rodriguez (2003b, p.129-137), the clusters are described as:

1. Social and academic language (4 items)

This cluster measures verbal precocity with four modes of language, reading, speaking, listening, and writing, in the native language. Gifted or potentially gifted bilingual Hispanic students were found to like to read, speak, listen, and write, and also achieve well in those areas.

2. Cultural sensitivity (3 items)

This cluster addresses the cultural aspect often lacking in other instruments, and associate an expressed and observable appreciation for the Hispanic culture with the Hispanic, ELL, gifted and potentially gifted students.

3. Familial (7 items)

This cluster considers the relevancy of family structures within the Hispanic culture and among these gifted and potentially gifted Hispanic, LEP children.

4. Motivation for learning (4 items)

This cluster includes demonstrated characteristic reflecting students' value for education and desire to succeed.

5. Collaboration (13 items)

This cluster focuses on students' abilities to lead and work with others in a cooperative nature.

6. Imagery (3 items)

This cluster is aligned with verbal precocity in Hispanic ELL students, and describes the characteristic exhibited by gifted and potentially gifted students who is able to imagine richly and describe vividly of events and stories.

7. Achievement (15 items)

This cluster reflects characteristics associated with academic giftedness perceived by teachers in a multi-facet way, including academic virtues and branching into more intrapersonal cognitive domains.

8. Support (5 items)

This cluster addresses the fact that students, who were perceived by the teachers as gifted and potentially gifted, still need support. Irby and Lara-Alecio used an analogy of 'the more bricks I have, the bigger building I can build", to deliver the idea that the stronger the support, the more likely the student is likely to succeed.

9. Creative performance (5 items)

This cluster measures students’ creative productivity in arts, music, drama, and physical activities.

10. Problem Solving (10 items)

This cluster contains items that deal regarding actions in solving problems, as well as cognitive functions of problem solving.

11. Locus of control (8 items)

This cluster is described as a loosely defined cluster encompassing the characteristics concerning the views of controlling factors one attributes to his/her own actions.

History of HBGSI

The HBGSI items originally were distilled from an extensive literature review in topics on gifted Hispanics, Hispanic familial/sociological/linguistic characteristics, Hispanic elementary children, and diverse gifted populations, including minority, rural, and urban (Irby, Lara-Alecio, & Rodriguez, 2003b). Over 400 characteristics were found, qualitatively coded, categorized, and eventually reduced to 90 items that composed the initial questionnaire. All items were worded as positive characteristics, and used a five-point scale. After a pilot study administered to bilingual teachers, results from descriptive statistics and agglomerative hierarchical cluster analysis showed that 78 of the items grouped into eleven clusters (Irby, Lara-Alecio, & Rodriguez, 2003b). One item (item 8 of the 78) was eventually deleted as further investigation suggested that this item had little or no added value to the instrument (Irby & Lara-Alecio, 2003).

The instrument initially came in a paper form, but has been digitized later. Now it utilizes an online service where teachers can log in, enter student scores and read student reports as well as their referral recommendations. Details of this online service are available at the website www.teachingbilingual.com.

Validation Studies of HBGSI

Initial evidence was collected from a sample of 61 elementary bilingual teachers who attended an annual state bilingual conference in Texas and volunteered to complete the 90-item questionnaire. The teachers taught various grades: Kindergarten (25%), first Grade (18%), second grade (30%), third grade (11%), and fourth grade (16%). Agglomerative hierarchical cluster analysis results suggested that the data be grouped into eleven clusters, and each cluster were found to have an alpha coefficients ranging from .62 to .91. Four tight clusters (Motivation for Learning, Social and Academic Language, Cultural Sensitivity and Imagery) and two moderately aligned clusters (Familial and Collaboration) were reported, all with alpha coefficients over .88.The HBGSI was shown to be effective in discriminating between students whose teachers would refer to gifted education testing and those they would not (Irby, Hernandez, Torres, & Gonzalez, 1997, as cited in Fultz, 2006).

With the 78-item HBGSI, Fultz (2004) studied 527 Hispanic students attending bilingual classroom programs from Kindergarten to 4th grade, in two schools in one ISD in Texas. Reliability coefficients for the HBGSI were found to range between .79 and .97, meeting the acceptable category range of reliability coefficients (Fultz, 2004). Fultz (2004) also explored the factor structure of HBGSI, using Principal Component Analysis

(PCA) with Varimax rotation. Evidence for five factors were provided, with item 8 being deleted from the instrument. Furthermore, Fultz(2004) was able to identify 75 students among the sampled students, who also were assessed in the Bilingual Verbal Ability Tests (BVAT) (Muñoz-Sandoval, Cummins, Alvarado, & Ruef, 1998). Evidence of concurrent validity ($r = .39$) coefficient when compared with the BVAT were provided (Fultz, 2004).

Concurrent validity was also studied with another sample of kindergarten students (Esquierdo, 2006). The 11 clusters and the total scores had a correlation with Naglieri Nonverbal Ability Test (NNAT) that ranged between .137 to .296 ($p<.01$). Concurrent validity of each cluster and the Woodcock Language Proficiency Batter-Revised (WLPB-R) were explored too. The analysis also found that there was a statistically significant difference in the students' performance on the WLPB-R subtests between the dichotomous groups of students identified and not identified as GT with the HBGSI.

Concurrent validity of each cluster in HBGSI was again explored in another study, with one particular subset of WLPB-R, *Verbal Analogies* in both English and Spanish as well as NNAT, for students from K to grade 3 (Contreras-Vanegas, 2011). Evidence of inter-rater reliability was also presented (Contreras-Vanegas, 2011).

Using a sample of 208 participants from K to grade 4, Irby and Lara-Alecio (1999) conducted exploratory factor analysis with the 78-item instrument and presented a shortened version of HBGSI with 38 items forming seven clusters, with a minimum of 2 items for each cluster. These items were extracted from eight out of the eleven original clusters, and all items in cluster *Cultural Sensitivity*, *Familial*, *Creative performance* were eliminated. One single item in the cluster of *Problem Solving* was retained and merged into the cluster of *Achievement*, while all other items in the cluster were eliminated.

Data Collection

Access was requested to use archived data in regard to HBGSI, collected throughout the ELLA project in elementary grades. Each student in ELLA project was assessed with HBGSI once a year, by his/her classroom teacher. Teachers attended a two and a half-hour training session, where the purpose of HBGSI and a comprehensive explanation of each component were provided, as well as a brief history of the development of the HBGSI and its importance in the screening stage for gifted and talented students. Furthermore, teachers received training on how to complete the HBGSI online along with a computer demonstration. Except for two teachers who refused to complete this instrument, all teachers were trained in this specific training session, and made plausible efforts completing this 77 item instrument for each participating student in their classrooms in the following 30-days timeframe.

Research Questions

The following questions guided my study:

1. What is the dimensionality of HBGSI and what is the degree of saturation by a general factor?

2. Based on the item response theory model, what are the properties of each item? Where on the latent trait continuum does each item provide the most discrimination among individuals?

3. Is it possible that a shortened version of HBGSI can be recommended, and what items will it include?

Data Analysis

Item-level data of 1117 samples of elementary students in grade 1 to grade 3 were be collected, coded, and analyzed. Analyses were performed using SPSS 17.0, MPlus 7.0 and R.

To answer question 1, exploratory factor analysis was conducted. Two types of exploratory factor analysis were considered and compared: (a) the traditional EFA method and (b) an exploratory form of bi-factor analysis where a general factor and a number of group factors are explored without a priori information.

To answer question 2, an item response theory model was established. The choice of the model, especially the dimension of the model, depended on the answer to question 1. Based on the findings, each item's properties including item difficulty, threshold and information curves were revealed.

To answer question 3, items that fit into the IRT model was selected. Selection criteria include item discrimination, item thresholds, item information, as well as its conceptual meaning. Items that satisfied these criteria best were retained in the shortened version of HBGSI.

CHAPTER IV

ANALYSIS OF DATA

Introduction

In this chapter, I present analysis conducted in response to the three research questions. I analyzed the structure of HBGSI, provide psychometric evidence based on Item response theory, and progressively worked towards an adapted and shortened version of HBGSI. This chapter is organized sequentially in regard to each research question.

Results

In the following, each research question is addressed with results presented in the same order as the questions were raised. Answers to earlier questions were referenced in answers to later questions.

Research Question 1

What is the dimensionality of HBGSI, and what is the degree of saturation by a general factor?

Initial evidence. Determining the dimension of an instrument is crucial in applying IRT because unidimensionality is a fundamental assumption of the more commonly used unidimensional IRT (UIRT) model. If a multidimensional instrument is fit improperly by UIRT models violating this assumption, there can be strong local dependence between certain items and item characteristics can be construed. To find out dimensionality of HBGSI, an exploratory factor analysis (EFA) was first conducted, treating item responses as categorical data. The scree plot is shown in Figure 2.

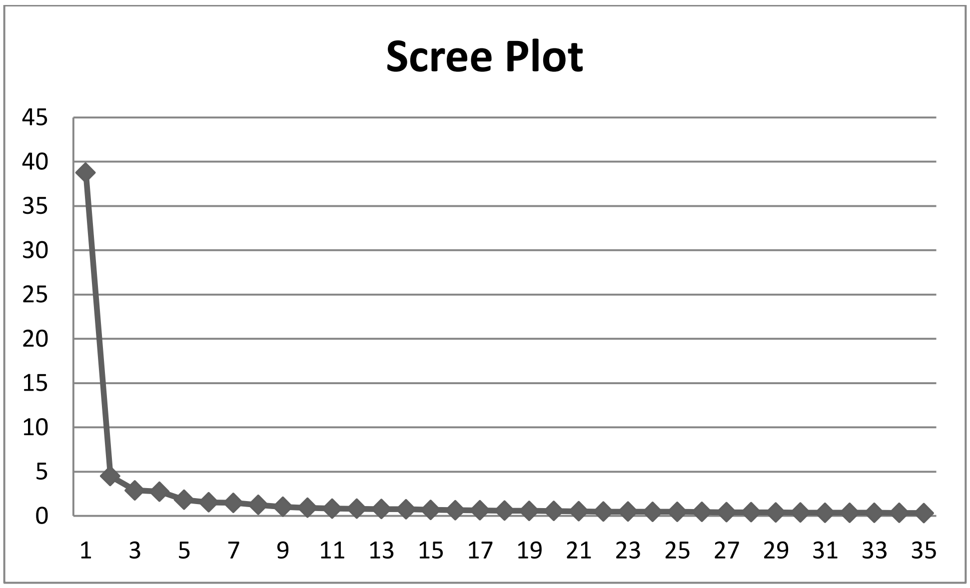

Figure 2. Scree plot

The first three eigenvalues were: 38.765, 4.485, 2.890. The ratio of first to second eigenvalue was large (8.6), well above the 3 to 1 ratio as evidence of a strong first factor. However, visual inspection of the scree plot was at least inclusive, because it could be argued that the ‘elbow’ area is around factor 9-11. Based on Kaiser’s rule (Kaiser, 1960), 9 factors had eigenvalue greater than 1, and should be counted as meaningful factors. Because K’s rule has been criticized and found to typically overestimate the number of components and sometimes underestimate (Zwick & Velicer, 1986), further MAP analysis (Velicer's,1976) has been conducted and revealed that 11 components exists (O'Conner, 2000). It does not seem that there is a gold standard rule of thumb for deciding when a response matrix is "unidimensional enough" for IRT modeling (Reise, Horan, & Blanchard, 2011), but multiple evidence seemed to point to a strong first factor, in a multidimensional structure. For the purpose of determining dimensionality for further IRT modeling, the key question is, as Reise et al. (2007) put it, do multiple dimensions that may emerge in a factor analysis interfere with our ability to implement a unidimensional IRT model? The focus is not the existence or nonexistence of multidimensionality, but if there is a strong a common factor and the strength of this factor. To answer these questions, bi-factor models were used and presented below.

Bifactor models. Many constructs in social science are broad ones comprised of several related domains, and can be represented by both second-order factor models and bifactor models. Compared to second-order factor models, bifactor models has not been common in literature and has not been widely used (Muthén & Asparouhov, 2012; Reise, Moore & Haviland, 2010), until recently. However, bifactor models seem to be getting increasing amount of attention and many applications in various domains are seen in 2010s (Biderman, 2013). In bi-factor models, each item has a nonzero loading on a general factor, and each of them also loads on at most one specific factor (a.k.a. group factor, domain factor) that represents the different sub-domains. The general factor accounts for the commonality of the items, and the specific factors accounts for the unique variances explained by the specific domains, with general factor being controlled. In comparison, in standard second-order models, items only have direct links to the lower-order, or first-order factors, which themselves are correlated. The first-order factors, in turn, then load on the higher-order, i.e., the second order factor, and form a hierarchical structure. Figure 3 illustrate the general differences between bi-factor models and second order factor models.

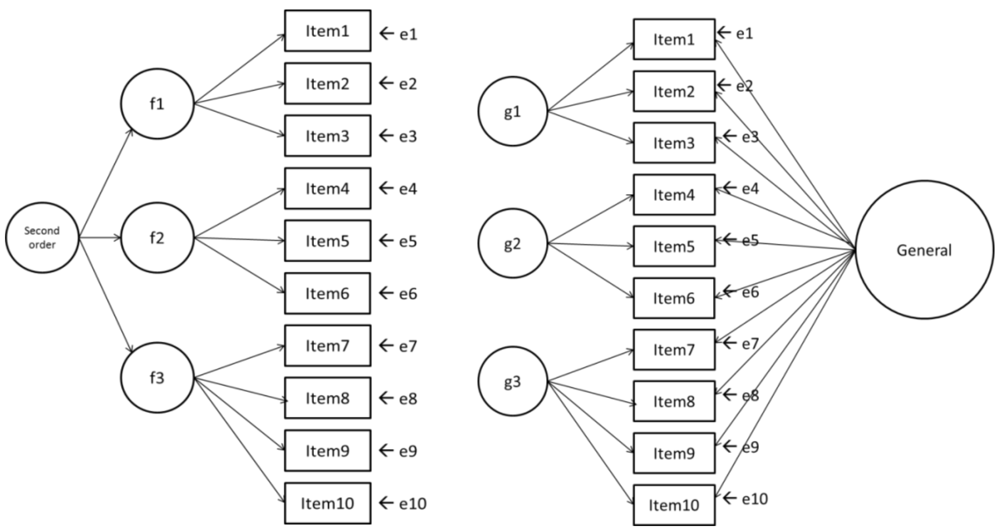

*Figure 3.*Second-order (left) and bifactor (right) models

Bi-factor models and second-order models are mathematically related. A bi-factor model can be equivalently mapped to a full second-order model, with added effects from the second-order factor directly to the items on top of a standard second-order model (Chen, West & Sousa, 2006). In other words, second-order models can be considered as a special case of bi-factor models, as it is a constrained case of the equivalent of bi-factor models. On a general basis, second-order models can be seen as actually nested in bi-factor models (Chen, et al., 2006; Yung, Thissen, & McLeod, 1999). Standard second order models are more restrictive than bi-factor models. Only in special cases where proportional constraints (Yung, et al., 1999) are imposed with Schmid-Leiman transformation method (Schmid & Leiman, 1957) are bi-factor models and second-order models equivalent.

Closely related mathematical representations lead to similar interpretations for second-order and bi-factor models (Chen et.al., 2006; Gustafsson & Balke, 1993). There is a correspondence between the general factor in the bi-factor model and the second-order factor in the second-order model. The domain specific factors respond to the uniqueness of the first-order factors in second-order model. General factor and domain factors are considered orthogonal in bi-factor models, which is in line with the second-order factor.

Chen et al. (2006) indicated that the bifactor model had several advantages over the second-order model when researchers are interested in both the domain specific factors and the general factor. In my study specifically, the bi-factor model was chosen for the following reasons:

1. Ease of interpretation. The structure of bi-factor models allowed us to examine a general factor, as well as domain specific factors with ease. Researchers in giftedness have shown the existence of g-factor (e.g., Carrol, 1993; Gottfredson, 2004; Jenson, 1998), as modeled in multiple hierarchical factor structures. Correspondingly, it can be speculated that a general factor well exists if a bi-factor model is fitted, and the general factor would correspond to the well-studied g-factor. The domain factors are

hypothesized to be independent of the general factors, which provide partial explanations to variance not accounted for in the general factor.

2. Interest in domain factors. Because Hispanic bilingual students are the focus of this study, it is of particular interest to study what factors are in play besides the g-factor. Bifactor models allow us to separate the sources of variances. The orthogonal structure allowed us to partition the items in meaningful groups without being clouded by the influence of the general factor. Specific domain knowledge can greatly help teachers better understand Hispanic bilingual students' behaviors and characteristics, and do a better job in identifying students who are gifted or potentially gifted.

3. Information on dimensionality. In IRT, it is crucial to determine the dimensionality of an instrument before models can be properly fitted. In bi-factor models we can examine how much of the item variance is due to the general construct, and how much is due to secondary dimensions. Item loadings on the general factor in the bi-factor model can be directly compared with loadings in a unidimensional model, providing clues as to how much distortions occurs if a multidimensional construct is fitted in a unidimensional model. Then examine the necessity and feasibility of creating a multidimensional IRT model or subscales can be examined (Reise, Morizot, & Hays, 2007).

An exploratory bi-factor model was first fitted to represent the item response structure. The polychoric correlation matrix was obtained and Schmid-Leiman orthogonization was performed, using the PSYC package in R. A different number of group factors were explored. Criteria for evaluating models include (a) the model be feasible (converge) and should have fit the data reasonably well (i.e. have acceptable fit statistics); (b) the model should be substantially interpretable---domains include items that are aligned in content; (c) the model should have no items or few items with cross loadings on the domain factors, and each domain should have at least 3 items. The five-group solution was the best option. In this bifactor structure, all items loaded on the general factor strongly, with average .64 (SD =.10, range .37 to .82). The five group factors extracted were substantially interpreted according to the items that have the highest loadings on them, as recommended by (Reise, et al., 2011). These five domain factors were thus named as Social Responsibility (SR); Academic Achievement (AA); Creative Performance (CP); Problem Solving (PS) and Native Language Proficiency (NLP). Loadings on the group factors were prevalently lower than on the general factor, except for Items 1 to 4, which loaded on the factor of native language proficiency. Table 2 shows detailed loadings. Only loadings with a magnitude greater than 0.20 were presented, with those greater than 0.30 in bold. Average loadings for each domain factor were: .40, .27, .34, .30, and .49 (counting only loadings greater than 0.20).

Table 2

Schmid-Leiman Exploratory Bifactor Analysis

Item	Cluster	Uni	Gen	SR	AA	CP	PS	NLP
Item 1	Social and Academic Language	**0.79**	**0.59**					**0.74**
Item 2	Social and Academic Language	**0.79**	**0.60**					**0.72**
Item 3	Social and Academic Language	**0.67**	**0.53**					**0.72**

Table 2 Continued

Item	Cluster	Uni	Gen	SR	AA	CP	PS	NLP
Item 4	Social and Academic Language	**0.68**	**0.55**					**0.68**
Item 5	Cultural Sensitivity	**0.74**	**0.65**	0.28				**0.30**
Item 6	Cultural Sensitivity	**0.73**	**0.64**	0.25			0.22	0.28
Item 7	Cultural Sensitivity	**0.72**	**0.65**	**0.31**				0.22
Item 8	Familial	**0.69**	**0.61**	**0.49**				
Item 9	Familial	**0.75**	**0.71**	**0.38**				
Item 10	Familial	**0.66**	**0.55**	**0.48**				
Item 11	Familial	**0.59**	**0.52**	**0.35**				
Item 12	Familial	**0.66**	**0.53**	**0.64**				
Item 13	Familial	**0.55**	**0.51**	**0.33**				
Item 14	Familial	**0.66**	**0.56**	**0.48**				
Item 15	Motivation	**0.80**	**0.72**	**0.51**				
Item 16	Motivation	**0.68**	**0.60**	**0.54**				
Item 17	Motivation	**0.85**	**0.74**	**0.51**				
Item 18	Motivation	**0.85**	**0.74**	**0.51**				
Item 19	Collaboration	**0.68**	**0.61**	**0.47**				
Item 20	Collaboration	**0.65**	**0.59**	**0.42**				
Item 21	Collaboration	**0.63**	**0.54**	**0.51**				
Item 22	Collaboration	**0.75**	**0.72**	0.24				
Item 23	Collaboration	**0.76**	**0.69**	**0.49**				
Item 24	Collaboration	**0.76**	**0.73**	0.28				
Item 25	Collaboration	**0.79**	**0.77**	**0.30**				
Item 26	Collaboration	**0.70**	**0.66**	**0.37**				
Item 27	Collaboration	**0.78**	**0.75**	**0.27**				
Item 28	Collaboration	**0.73**	**0.70**	**0.37**				
Item 29	Collaboration	**0.65**	**0.57**	**0.51**				
Item 30	Collaboration	**0.73**	**0.69**	**0.34**				
Item 31	Collaboration	**0.69**	**0.68**	0.24				
Item 32	Imagery	**0.78**	**0.75**			0.26		
Item 33	Imagery	**0.76**	**0.71**			**0.30**		
Item 34	Imagery	**0.78**	**0.75**		0.20	0.20		
Item 35	Achievement	**0.81**	**0.81**		**0.31**			
Item 36	Achievement	**0.83**	**0.83**		**0.36**			
Item 37	Achievement	**0.79**	**0.80**		**0.35**			
Item 38	Achievement	**0.78**	**0.77**		0.23	0.21		
Item 39	Achievement	**0.72**	**0.73**		**0.34**			

Table 2 Continued

Item	Cluster	Uni	Gen	SR	AA	CP	PS	NLP
Item 40	Achievement	**0.76**	**0.77**		**0.35**			
Item 41	Achievement	**0.62**	**0.62**		0.23			
Item 42	Achievement	**0.81**	**0.81**		**0.32**			
Item 43	Achievement	**0.80**	**0.80**		0.27			
Item 44	Achievement	**0.76**	**0.75**		0.21	0.26		
Item 45	Achievement	**0.57**	**0.56**					0.24
Item 46	Achievement	**0.72**	**0.72**		0.26			
Item 47	Achievement	**0.65**	**0.63**			0.27		
Item 48	Achievement	**0.70**	**0.70**		0.27	0.23		
Item 49	Achievement	**0.76**	**0.74**		0.21	0.21		
Item 50	Support	**0.55**	**0.48**	0.31			0.29	
Item 51	Support	**0.44**	**0.41**			0.23	0.35	
Item 52	Support	**0.38**	**0.37**			0.23	0.24	
Item 53	Support	**0.58**	**0.59**		0.27			
Item 54	Support	**0.57**	**0.51**				0.27	0.53
Item 55	Creative Performance	**0.74**	**0.62**			**0.64**		
Item 56	Creative Performance	**0.78**	**0.64**			**0.65**		
Item 57	Creative Performance	**0.73**	**0.63**			**0.62**		
Item 58	Creative Performance	**0.61**	**0.55**			**0.56**		
Item 59	Creative Performance	**0.56**	**0.50**			**0.31**	**0.41**	
Item 60	Problem Solving	**0.64**	**0.63**		0.22		**0.34**	
Item 61	Problem Solving	**0.70**	**0.71**		**0.32**			
Item 62	Problem Solving	**0.77**	**0.76**		0.24			
Item 63	Problem Solving	**0.65**	**0.64**		0.20			
Item 64	Problem Solving	**0.78**	**0.78**		0.23	0.20		
Item 65	Problem Solving	**0.60**	**0.55**			**0.50**		
Item 66	Problem Solving	**0.55**	**0.54**			0.26		
Item 67	Problem Solving	**0.55**	**0.52**			**0.33**		
Item 68	Problem Solving	**0.60**	**0.55**				**0.43**	
Item 69	Problem Solving	**0.64**	**0.61**				**0.39**	
Item 70	Locus of Control	**0.77**	**0.78**		0.28			
Item 71	Locus of Control	**0.64**	**0.63**		0.20		0.24	
Item 72	Locus of Control	**0.73**	**0.72**		0.21		0.24	
Item 73	Locus of Control	**0.75**	**0.76**		**0.31**			
Item 74	Locus of Control	**0.55**	**0.48**				**0.47**	
Item 75	Locus of Control	**0.58**	**0.54**	**0.41**				

Table 2 Continued

Item	Cluster	Uni	Gen	SR	AA	CP	PS	NLP
Item 76	Locus of Control	**0.64**	**0.59**	**0.35**			0.23	
Item 77	Locus of Control	**0.69**	**0.61**	**0.50**				

Eigenvalues of the general and five group factors were respectively: 32.8, 5.6, 2.2, 3.2, 2.1, 2.9, with a great first-to-second eigenvalue ratio. Across all items, the average common variance explained by general factor was 66.9%. Omega general for total score was 0.83. These evidence seemed to suggest that the general factor is a strong, but not dominant.

In order to evaluate the dominance of the general factor, I followed recommendations by Reise et al. (2011), and compared item loadings in a unidimensionamodel, and in a bi-factor model. Items in HBGSI encompassed varied degree of disparity between loadings in unidimensional models and in bifactor models, as shown in Table 2. While items in certain clusters (Achievement, Problem Solving, and Locus of Control) have minimum differences that are hardly appreciable, for items in other clusters (Social and Academic Language, Cultural Sensitivity, Familial, Motivation) the difference was larger. When multidimensional data is fit to one-factor models, factor loadings could possibly be distorted because they were in fact influenced by group factors that were not accounted for. In cases where a subset of items share common variance due to both the general factor and a group factor, if modeled as unidimensional, the single factor likely embodies both the true general factor and some of the group factor to certain extents. This single factor in a unidimensional model may be "pulled towards these highly correlated items" (Reise et al., 2007, p25). It also means that the latent construct modeled by the unidimensional model might be distorted from the true general construct underlying all items. When the general factor is dominant enough, UIRT can be legitimately fit to data despite existence of multidimensionality (Reise et al., 2007). However, in my study, prevalent existence of between-model changes in item loadings in certain clusters barred such possibility and forbade the attempt to fit all items in HBGSI in a unidimensional IRT model. The conclusion was also supported by the fact that the unidimensional model had poor model fit with RMSEA = 0.096 and CFI = 0.23, suggesting that one single construct was not enough to explain all items in the instrument of HBGSI.

Conclusion. I concluded that there was one strong factor in HBGSI, but it was not dominant enough for application of a single UIRT model for this instrument. Instead, a bifactor model revealed that a strong general factor and five group factors, which can explain the structure well both statistically and substantially.

Research Question 2

Based on the IRT model, what are the properties of each item?

Aided by clues given by the exploratory bifactor analysis, a compensatory multidimensional IRT based model for the entire instrument of HBGSI was built. Each item was considered to be influenced by two dimensions: the first dimension that represented general giftedness, and a second dimension which represented one of the five subdomains: Academic Achievement (AA); Social Responsibility (SR); Creative Performance (CP); Problem Solving (PS) and Native Language Proficiency (NLP). The group factor for each item was chosen as the one where the item has the highest magnitude of loading in the bifactor exploratory model. Once the MIRT model is built, it was found that some items have numerically small and statistically nonsignificant (i.e., not significantly different from zero) discrimination parameter in regard to the group factor. These items were then set to have influence only by the general construct. This iterative verification process was employed until the final model can be established with all the discrimination parameters being significantly different from zero.

The MIRT model was evaluated in MPlus 7.0 (Muthén & Muthén, 1998-2012) with the WLSMV estimator. A robust weighted least squares estimator using a diagonal weight matrix with standard errors and mean- and variance- adjusted chi-square test statistic that use a full weight matrix were deployed by specifying WLSMV (Muthén & Muthén, 1998-2012) as the estimator, and greatly improves computational speed as there are many dimensions of integration and categorical outcomes in the MIRT analysis. With WLSMV, Chi-square statistics was not provided. This model had CFI = .92 and RMSEA=.06, suggesting acceptable model fit. Model fit would likely improve if items with cross loadings were allowed, i.e., if items were allowed to be influenced by more than one domain affiliation. Item parameters are presented in Table 3. The slopes are expressed in the more commonly used logistic metric with scaling of D =1.7 applied.

Table 3

Discrimination Parameters

		Discrimination Parameters		
Item	Cluster	General	Domain	Domain Name
Item_01	Social and Academic Language	1.03 (0.04)	1.25 (0.03)	NLP

Table 3 Continued

Item	Cluster	Discrimination Parameter: General	Discrimination Parameter: Domain	Domain Name
Item_02	Social and Academic Language	1.05 (0.03)	1.23 (0.03)	NLP
Item_03	Social and Academic Language	0.92 (0.04)	1.26 (0.03)	NLP
Item_04	Social and Academic Language	0.98 (0.04)	1.18 (0.03)	NLP
Item_05	Cultural Sensitivity	1.29 (0.03)	0.67 (0.04)	NLP
Item_06	Cultural Sensitivity	1.28 (0.03)	0.66 (0.03)	NLP
Item_07	Cultural Sensitivity	1.27 (0.03)	0.25 (0.04)	SR
Item_08	Familial	1.11 (0.03)	0.71 (0.04)	SR
Item_09	Familial	1.27 (0.03)	0.56 (0.03)	SR
Item_10	Familial	0.95 (0.04)	1.00 (0.04)	SR
Item_11	Familial	0.91 (0.04)	0.75 (0.04)	SR
Item_12	Familial	0.90 (0.04)	1.11 (0.04)	SR
Item_13	Familial	0.90 (0.04)	0.56 (0.04)	SR
Item_14	Familial	0.96 (0.04)	1.03 (0.03)	SR
Item_15	Motivation	1.26 (0.03)	0.81 (0.03)	SR
Item_16	Motivation	1.03 (0.04)	0.87 (0.04)	SR
Item_17	Motivation	1.29 (0.03)	0.92 (0.03)	SR
Item_18	Motivation	1.29 (0.03)	0.94 (0.03)	SR
Item_19	Collaboration	1.07 (0.04)	0.74 (0.04)	SR
Item_20	Collaboration	1.06 (0.04)	0.64 (0.04)	SR
Item_21	Collaboration	0.95 (0.04)	0.87 (0.04)	SR
Item_22	Collaboration	1.32 (0.02)	0.30 (0.04)	SR
Item_23	Collaboration	1.21 (0.03)	0.77 (0.03)	SR
Item_24	Collaboration	1.33 (0.02)	0.34 (0.04)	SR
Item_25	Collaboration	1.38 (0.02)	0.37 (0.04)	SR
Item_26	Collaboration	1.19 (0.03)	0.49 (0.04)	SR
Item_27	Collaboration	1.36 (0.02)	0.34 (0.04)	SR
Item_28	Collaboration	1.24 (0.03)	0.51 (0.04)	SR
Item_29	Collaboration	1.01 (0.04)	0.80 (0.04)	SR
Item_30	Collaboration	1.25 (0.02)	0.44 (0.04)	SR
Item_31	Collaboration	1.23 (0.03)	0.27 (0.04)	SR
Item_32	Imagery	1.35 (0.02)	0.61 (0.03)	CP

Table 3 Continued

Item	Cluster	Discrimination Parameter General	Discrimination Parameter Domain	Domain Name
Item_33	Imagery	1.29 (0.02)	0.71 (0.03)	CP
Item_34	Imagery	1.36 (0.02)	0.49 (0.03)	CP
Item_35	Achievement	1.43 (0.02)	0.46 (0.04)	AA
Item_36	Achievement	1.43 (0.02)	0.63 (0.03)	AA
Item_37	Achievement	1.40 (0.02)	0.47 (0.04)	AA
Item_38	Achievement	1.40 (0.02)	0.20 (0.04)	AA
Item_39	Achievement	1.25 (0.03)	0.62 (0.04)	AA
Item_40	Achievement	1.30 (0.02)	0.75 (0.04)	AA
Item_41	Achievement	1.13 (0.03)	0.10 (0.05)	AA
Item_42	Achievement	1.44 (0.02)	0.34 (0.03)	AA
Item_43	Achievement	1.45 (0.02)	0.19 (0.04)	AA
Item_44	Achievement	1.37 (0.02)	0.24 (0.03)	CP
Item_45	Achievement	1.02 (0.04)	0.32 (0.04)	NLP
Item_46	Achievement	1.31 (0.02)	0.19 (0.04)	AA
Item_47	Achievement	1.16 (0.03)	0.23 (0.04)	CP
Item_48	Achievement	1.28 (0.02)	0.11 (0.04)	AA
Item_49	Achievement	1.37 (0.02)	0.16 (0.03)	CP
Item_50	Support	0.94 (0.04)	0.39 (0.04)	SR
Item_51	Support	0.78 (0.04)	0.85 (0.05)	PS
Item_52	Support	0.68 (0.04)	0.66 (0.05)	PS
Item_53	Support	1.02 (0.03)	0.48 (0.05)	AA
Item_54	Support	0.95 (0.04)	0.80 (0.04)	NLP
Item_55	Creative Performance	1.11 (0.03)	1.06 (0.03)	CP
Item_56	Creative Performance	1.17 (0.03)	1.09 (0.03)	CP
Item_57	Creative Performance	1.13 (0.03)	1.03 (0.03)	CP
Item_58	Creative Performance	0.98 (0.04)	0.94 (0.03)	CP
Item_59	Creative Performance	1.00 (0.03)	0.72 (0.04)	PS
Item_60	Problem Solving	1.15 (0.03)	0.52 (0.04)	PS
Item_61	Problem Solving	1.21 (0.03)	0.67 (0.04)	AA
Item_62	Problem Solving	1.40 (0.02)	0.15 (0.04)	AA
Item_63	Problem Solving	1.20 (0.03)	N/A	N/A
Item_64	Problem Solving	1.43 (0.02)	N/A	N/A
Item_65	Problem Solving	0.97 (0.04)	0.86 (0.03)	CP
Item_66	Problem Solving	0.97 (0.04)	0.35 (0.04)	CP

Table 3 Continued

		Discrimination Parameter		
Item	Cluster	General	Domain	Domain Name
Item_67	Problem Solving	0.95 (0.04)	0.48 (0.04)	CP
Item_68	Problem Solving	1.07 (0.03)	0.67 (0.05)	PS
Item_69	Problem Solving	1.14 (0.03)	0.80 (0.04)	PS
Item_70	Locus of Control	1.40 (0.02)	0.23 (0.04)	AA
Item_71	Locus of Control	1.18 (0.03)	0.15 (0.05)	PS
Item_72	Locus of Control	1.32 (0.02)	0.33 (0.04)	PS
Item_73	Locus of Control	1.33 (0.02)	0.42 (0.04)	AA
Item_74	Locus of Control	0.98 (0.04)	0.16 (0.04)	SR
Item_75	Locus of Control	0.93 (0.04)	0.65 (0.04)	SR
Item_76	Locus of Control	1.05 (0.04)	0.60 (0.04)	SR
Item_77	Locus of Control	1.07 (0.03)	0.85 (0.04)	SR

The slopes in general dimensions were substantial to high, and were low in domain dimensions. Both the general and domain constructs were deemed important in HBGSI. Domain constructs provide valuable insights in regard to the variations not accounted by the general construct, and complement the general construct that describes the overall giftedness in bilingual Hispanic students defined in a broad sense. In general, the five domain constructs influenced items to a lesser degree than the general item, but there were still items with moderate discriminating power for each and every one of these domains respectively. These items complemented the general giftedness construct and provided a more comprehensive picture in depicting and measuring giftedness. Items in domain Native Language Proficiency (NLP) had the highest discriminating power, while items in domain Academic Achievement (AA) had the least. This is explainable because compared to items for domain NLP, items in domain AA were more discriminant in regard to general giftedness. From the item factor analysis view, more variance of these items was explained by the general giftedness factor, thus less by the domain factors.

Thresholds. In IRT, thresholds depict the location where responses are likely to fall in a particular category or higher. Four threshold values were reported for this five category scale, and listed item-wise in Table 4.

Table 4

Thresholds

Item	Cluster	Threshold 1	Threshold 2	Threshold 3	Threshold 4
Item_01	Social and Academic Language	-0.97 (0.05)	-0.50 (0.04)	-0.01 (0.04)	0.54 (0.04)
Item_02	Social and Academic Language	-0.84 (0.04)	-0.34 (0.04)	0.18 (0.04)	0.71 (0.04)
Item_03	Social and Academic Language	-1.45 (0.06)	-1.00 (0.05)	-0.32 (0.04)	0.25 (0.04)
Item_04	Social and Academic Language	-1.48 (0.06)	-1.03 (0.05)	-0.32 (0.04)	0.21 (0.04)
Item_05	Cultural Sensitivity	-1.88 (0.08)	-1.26 (0.05)	-0.30 (0.04)	0.44 (0.04)
Item_06	Cultural Sensitivity	-1.67 (0.06)	-1.06 (0.05)	-0.11 (0.04)	0.56 (0.04)
Item_07	Cultural Sensitivity	-1.87 (0.08)	-1.26 (0.05)	-0.29 (0.04)	0.35 (0.04)
Item_08	Familial	-1.83 (0.07)	-1.08 (0.05)	-0.33 (0.04)	0.26 (0.04)
Item_09	Familial	-1.81 (0.07)	-1.16 (0.05)	-0.28 (0.04)	0.37 (0.04)
Item_10	Familial	-1.80 (0.07)	-1.33 (0.05)	-0.67 (0.04)	-0.07 (0.04)
Item_11	Familial	-1.37 (0.05)	-0.94 (0.04)	-0.28 (0.04)	0.23 (0.04)
Item_12	Familial	-1.94 (0.08)	-1.43 (0.06)	-0.79 (0.04)	-0.20 (0.04)
Item_13	Familial	-1.17 (0.05)	-0.47 (0.04)	0.14 (0.04)	0.61 (0.04)
Item_14	Familial	-1.82 (0.07)	-1.18 (0.05)	-0.52 (0.04)	0.12 (0.04)
Item_15	Motivation	-1.74 (0.07)	-1.18 (0.05)	-0.33 (0.04)	0.30 (0.04)
Item_16	Motivation	-2.21 (0.10)	-1.49 (0.06)	-0.92 (0.04)	-0.24 (0.04)
Item_17	Motivation	-1.70 (0.07)	-1.11 (0.05)	-0.29 (0.04)	0.27 (0.04)
Item_18	Motivation	-1.63 (0.06)	-1.10 (0.05)	-0.31 (0.04)	0.25 (0.04)

Table 4 Continued

Item	Cluster	Threshold 1	Threshold 2	Threshold 3	Threshold 4
Item_19	Collaboration	-2.36 (0.12)	-1.81 (0.07)	-0.89 (0.04)	-0.19 (0.04)
Item_20	Collaboration	-2.24 (0.10)	-1.43 (0.06)	-0.46 (0.04)	0.26 (0.04)
Item_21	Collaboration	-1.70 (0.07)	-1.05 (0.05)	-0.27 (0.04)	0.36 (0.04)
Item_22	Collaboration	-1.40 (0.06)	-0.67 (0.04)	0.09 (0.04)	0.65 (0.04)
Item_23	Collaboration	-1.88 (0.08)	-1.17 (0.05)	-0.32 (0.04)	0.35 (0.04)
Item_24	Collaboration	-1.49 (0.06)	-0.66 (0.04)	0.15 (0.04)	0.72 (0.04)
Item_25	Collaboration	-1.41 (0.06)	-0.62 (0.04)	0.23 (0.04)	0.80 (0.04)
Item_26	Collaboration	-1.94 (0.08)	-1.14 (0.05)	-0.23 (0.04)	0.41 (0.04)
Item_27	Collaboration	-1.72 (0.07)	-0.99 (0.05)	0.04 (0.04)	0.65 (0.04)
Item_28	Collaboration	-1.60 (0.06)	-0.79 (0.04)	-0.05 (0.04)	0.55 (0.04)
Item_29	Collaboration	-1.90 (0.08)	-1.24 (0.05)	-0.45 (0.04)	0.27 (0.04)
Item_30	Collaboration	-1.49 (0.06)	-0.79 (0.04)	0.14 (0.04)	0.69 (0.04)
Item_31	Collaboration	-1.68 (0.07)	-0.97 (0.05)	-0.15 (0.04)	0.48 (0.04)
Item_32	Imagery	-1.40 (0.06)	-0.71 (0.04)	0.14 (0.04)	0.90 (0.04)
Item_33	Imagery	-1.33 (0.05)	-0.63 (0.04)	0.14 (0.04)	0.76 (0.04)
Item_34	Imagery	-1.06 (0.05)	-0.40 (0.04)	0.45 (0.04)	1.04 (0.05)
Item_35	Achievement	-1.76 (0.07)	-0.95 (0.05)	0.09 (0.04)	0.75 (0.04)
Item_36	Achievement	-1.90 (0.08)	-1.04 (0.05)	-0.06 (0.04)	0.64 (0.04)
Item_37	Achievement	-1.41 (0.06)	-0.64 (0.04)	0.32 (0.04)	0.86 (0.04)
Item_38	Achievement	-1.34 (0.05)	-0.53 (0.04)	0.37 (0.04)	0.91 (0.04)
Item_39	Achievement	-1.72 (0.07)	-0.98 (0.05)	-0.07 (0.04)	0.60 (0.04)
Item_40	Achievement	-1.31 (0.05)	-0.63 (0.04)	0.07 (0.04)	0.68 (0.04)
Item_41	Achievement	-1.31 (0.05)	-0.51 (0.04)	0.25 (0.04)	0.79 (0.04)
Item_42	Achievement	-1.67 (0.07)	-0.89 (0.04)	0.10 (0.04)	0.72 (0.04)
Item_43	Achievement	-1.43 (0.06)	-0.68 (0.04)	0.07 (0.04)	0.66 (0.04)
Item_44	Achievement	-1.01 (0.05)	-0.34 (0.04)	0.41 (0.04)	0.94 (0.04)
Item_45	Achievement	-1.36 (0.05)	-0.79 (0.04)	-0.11 (0.04)	0.60 (0.04)
Item_46	Achievement	-1.61 (0.06)	-0.74 (0.04)	0.08 (0.04)	0.62 (0.04)
Item_47	Achievement	-1.47 (0.06)	-0.77 (0.04)	0.11 (0.04)	0.70 (0.04)
Item_48	Achievement	-1.55 (0.06)	-0.73 (0.04)	0.17 (0.04)	0.77 (0.04)
Item_49	Achievement	-1.24 (0.05)	-0.51 (0.04)	0.27 (0.04)	0.81 (0.04)
Item_50	Support	-2.04 (0.09)	-1.56 (0.06)	-0.73 (0.04)	-0.05 (0.04)
Item_51	Support	-1.24 (0.05)	-0.80 (0.04)	0.20 (0.04)	0.70 (0.04)
Item_52	Support	-1.26 (0.05)	-0.50 (0.04)	0.51 (0.04)	1.05 (0.05)
Item_53	Support	-1.33 (0.05)	-0.66 (0.04)	0.11 (0.04)	0.71 (0.04)
Item_54	Support	-1.31 (0.05)	-0.72 (0.04)	-0.03 (0.04)	0.53 (0.04)
Item_55	Creative Performance	-1.11 (0.05)	-0.43 (0.04)	0.38 (0.04)	1.00 (0.05)

Table 4 Continued

Item	Cluster	Threshold 1	Threshold 2	Threshold 3	Threshold 4
Item_56	Creative Performance	-1.02 (0.05)	-0.40 (0.04)	0.51 (0.04)	1.08 (0.05)
Item_57	Creative Performance	-1.20 (0.05)	-0.44 (0.04)	0.50 (0.04)	0.97 (0.05)
Item_58	Creative Performance	-0.60 (0.04)	0.12 (0.04)	0.92 (0.04)	1.39 (0.05)
Item_59	Creative Performance	-1.44 (0.06)	-0.71 (0.04)	0.19 (0.04)	0.78 (0.04)
Item_60	Problem Solving	-1.54 (0.06)	-0.92 (0.04)	0.12 (0.04)	0.79 (0.04)
Item_61	Problem Solving	-1.61 (0.06)	-0.86 (0.04)	0.00 (0.04)	0.72 (0.04)
Item_62	Problem Solving	-1.41 (0.06)	-0.59 (0.04)	0.24 (0.04)	0.77 (0.04)
Item_63	Problem Solving	-1.37 (0.05)	-0.56 (0.04)	0.19 (0.04)	0.73 (0.04)
Item_64	Problem Solving	-1.43 (0.06)	-0.72 (0.04)	0.28 (0.04)	0.87 (0.04)
Item_65	Problem Solving	-0.72 (0.04)	0.02 (0.04)	0.90 (0.04)	1.39 (0.05)
Item_66	Problem Solving	-1.77 (0.07)	-1.01 (0.05)	-0.14 (0.04)	0.54 (0.04)
Item_67	Problem Solving	-0.93 (0.04)	-0.36 (0.04)	0.48 (0.04)	1.00 (0.05)
Item_68	Problem Solving	-1.80 (0.07)	-0.94 (0.04)	0.10 (0.04)	0.75 (0.04)
Item_69	Problem Solving	-1.54 (0.06)	-0.86 (0.04)	0.34 (0.04)	0.89 (0.04)
Item_70	Locus of Control	-1.61 (0.06)	-1.06 (0.05)	-0.18 (0.04)	0.50 (0.04)
Item_71	Locus of Control	-1.71 (0.07)	-1.08 (0.05)	-0.17 (0.04)	0.52 (0.04)
Item_72	Locus of Control	-1.55 (0.06)	-0.78 (0.04)	0.25 (0.04)	0.85 (0.04)
Item_73	Locus of Control	-1.36 (0.05)	-0.69 (0.04)	0.09 (0.04)	0.66 (0.04)
Item_74	Locus of Control	-1.76 (0.07)	-1.04 (0.05)	-0.24 (0.04)	0.45 (0.04)
Item_75	Locus of Control	-1.63 (0.06)	-1.08 (0.05)	-0.41 (0.04)	0.12 (0.04)
Item_76	Locus of Control	-1.83 (0.07)	-1.25 (0.05)	-0.57 (0.04)	0.02 (0.04)
Item_77	Locus of Control	-1.61 (0.06)	-1.09 (0.05)	-0.54 (0.04)	0.06 (0.04)

Thirty six items among all 77 items had negative third thresholds. Among them, five items had negative fourth thresholds (Items 10, 12, 16, 19, 50). Take Item 16 as an example, the fourth threshold being -.235 meant that students who were .235 standard deviation below average had 50-50 chance in in scoring the highest category in this item. In other words, all students who performed above average were more likely than not to score the same points (5 points) for this item. This is substantially understandable, because it is expected that all students, gifted or not, attend school on a regular basis, and thus a large number of student would 'always' exhibit such characteristic. Statistically, it indicated that this item lacked discrimination power among students in the top half because it was hard to tell apart the best students from the average with all of them gaining the same score. Such items are likely better items in diagnosing students who are at risk, rather than students who perform at the higher end.

General giftedness. Similar to results in the exploratory bifactor model, discriminant parameter for the general factor is strong across all items. The average discriminate parameter is 1.16 (SD = 0.18, range = 0.68 to 1.44). Parameters were scaled up to fit the commonly used logistic model parameters. Cluster wise, items in the Familial and Support cluster seemed to have the least discriminating power in regard to the general factor, when compared to items in other clusters. In contrast, items in the cluster Achievement and Cultural Sensitivity are generally most discriminant. Cluster Problem Solving and Locus of Control had a mixture of both.

Domain constructs. In general, the five domain constructs influenced the items to a lesser degree than the general construct, but there were still items with moderate discriminating power for each and every one of these domains respectively. These items complemented the general giftedness construct and provided a more comprehensive picture in depicting and measuring giftedness. Item in domain NLP had the highest domain discrimination, while items in domain AA had the least. This is explainable because compared to items in NLP, items in AA were more discriminant in regard to general giftedness.

Research Question 3

Based on the IRT model, is it possible that a shortened version of HBGSI can be recommended, and what items will it include?

As a teacher rating scale, HBGSI asks teachers to not only know each student's daily behavior well, but also to familiarize themselves with the items. Teachers need to be able and willing to carefully evaluate each student against each and every item in the instrument. If among the collections of 77 items, items that best measure Hispanic bilingual giftedness can be selected, then teachers will have the option of a shorter version of HBGSI, a potential relief to their time and energy.

Selection. To select the best items, I first looked at item locations. Because HBGSI is intended for use in gifted screening, items located at the higher end are most desirable. Five items (Items 10, 12, 16, 19, 50) were excluded based on this criteria, because their highest threshold were below zero, suggesting that even students below average has a 50% chance to score the highest category. These items would thus not contribute much in differentiating the highly gifted students, who perform better than the average.

Because both the general and domain constructs are of concern in HBGSI, items that can discriminate well at both constructs are most desirable. Initially, items were selected if they were in the upper 50th percentile in regard to the discrimination parameter in general factor as well as their respective domains. Only 12 items satisfied these criteria, and they are listed as follows in Table 5.As Table 5 shows, domains of *Problem Solving* and *Native Language Proficiency* were not represented in the above collection, due to the fact that top items in these two domains did not have high discriminating power in regard to the general factor. Still, these domains included important characteristics of HBGSI in developing a comprehensive depiction of what giftedness is in Hispanic Bilingual students. An adaptation that let go of such characteristics would sacrifice the broad spectrum of this instrument, and take away from the inclusive goal embraced in the original development of HBGSI. Thus, a decision was made to retain items in the top third in regard to their discriminating power, for either the general or respective domain constructs.

Table 5

Initial selection

Item	Description	Domain
Item_15	Values education; sees education as a way to improve status	SR
Item_17	Appears to have sustained motivation to succeed; is persistent	SR
Item_18	Is motivated to learn; exhibits a desire for learning	SR
Item_23	Possesses leadership qualities in relation to working in the peer group; works well with others	SR
Item_33	Exhibits language (speaking) rich in imagery	CP
Item_35	Has ability to generalize learning to other areas and show relationships among apparently unrelated ideas	AA
Item_36	Has the ability to use stored knowledge to solve problems	AA
Item_37	Reasons by analog or contrast	AA
Item_39	The relationship between learning and language is consistent in the areas of math and science; level of competency is equal in all of those areas	AA
Item_40	Performs at or above grade level in math; has high math abilities; likes to do math problems	AA
Item_61	Performs at or above grade level in science; likes to do science experiments	AA
Item_73	Has effective test taking skills	AA

After adopting the new criteria, a total of 40 items were retained. These items are listed in Table 6 as follows.

Table 6

Items Selected with New Criteria

Item	Description	Domain
Item_01	Likes to read in native language; is a proficient reader in native language	NLP
Item_03	Likes to speak in native language; is a proficient speaker in native language	NLP
Item_05	Aware of own language/culture and takes pride in language/culture	NLP
Item_14	Has strong emotional support from families	SR
Item_15	Values education; sees education as a way to improve status	SR
Item_17	Appears to have sustained motivation to succeed; is persistent	SR
Item_18	Is motivated to learn; exhibits a desire for learning	SR
Item_21	Indirect in giving criticism; avoid conflict; is amenable	SR
Item_22	Possesses leadership qualities in relation to working in the peer group; works well with others	SR
Item_24	Demonstrates ability for giving advice and judgments in disputes and in planning strategies	SR
Item_25	Effective at setting goals	SR
Item_27	Is able to evaluate events and people	SR
Item_29	Likes to please; sensitivity to the opinions of others	SR
Item_32	Exhibits language (speaking) rich in imagery	CP
Item_33	Is imaginative in story telling	CP
Item_34	Exhibits language (writing) rich in imagery	CP
Item_35	Has ability to generalize learning to other areas and show relationships among apparently unrelated ideas	AA
Item_36	Has the ability to use stored knowledge to solve problems	AA
Item_37	Reasons by analog or contrast	AA
Item_38	Talents demonstrated through various projects and interests at home or in the community	AA

Table 6 Continued

Item	Description	Domain
Item_39	The relationship between learning and language is consistent in the areas of math and science; level of competency is equal in all of those areas	AA
Item_40	Performs at or above grade level in math; has high math abilities; likes to do math problems	AA
Item_42	Perceives cause and effect relationships	AA
Item_43	Is self-directed in activities and is methodological	AA
Item_44	Has an entrepreneurial ability/spirit	CP
Item_46	Is curious; always investigating or asking questions; eager to try out new ideas; likes to take risk	AA
Item_49	Uses intuition	CP
Item_51	Prefers alternative assessments rather than standardized or teacher made tests	PS
Item_53	Needs minimal support in second language acquisition	AA
Item_55	Exhibits creativity in movement, dance and other physical activities	CP
Item_56	His/her performance in arts, music, dance, when using his/her native culture is unique	CP
Item_57	Adept in visual and/or performing arts; is talented in art, music or drama	CP
Item_58	Creative in lyric production to songs; likes to compose poetry	CP
Item_61	Performs at or above grade level in science; likes to do science experiments	AA
Item_62	Appears to have more patience in dealing with tasks not easily resolved	AA
Item_64	Exhibits high nonverbal fluency and originality	N/A
Item_69	Performs better on spatial fluency tasks as opposed to verbal fluency tasks (fluency is designed as the generation of many ideas - may be verbal on nonverbal)	PS
Item_70	Exhibits steadfast self-concept and self-confidence	AA
Item_72	Has the ability to meaningfully manipulate symbolism in his/her culture	PS
Item_73	Has effective test taking skills	AA

Validation

Information curves and standard error. One distinguished feature of Item Response Theory models is that items provide varied amount of information at different levels of latent trait. Where information is high, standard error is low and the item/test is highly reliable.

As shown in Figure 4, standard errors were at a minimum where latent trait was close to zero across all items in regard to the group and domain constructs, meaning these items provided maximum information to the group of students performing at or around average. These traits fit the screening purpose as a screening instrument used in the first phase of identification. It is important that a greater number of students are recommended for further testing, than the number of students who will be eventually admitted in gifted programs. In practice, typically the top half of students were recommended. Thus items that gave more information for students around the group of student with average ability were desirable for HBGSI. Among the 40 items retained thus far, no items were eliminated based on the location where information peaked.

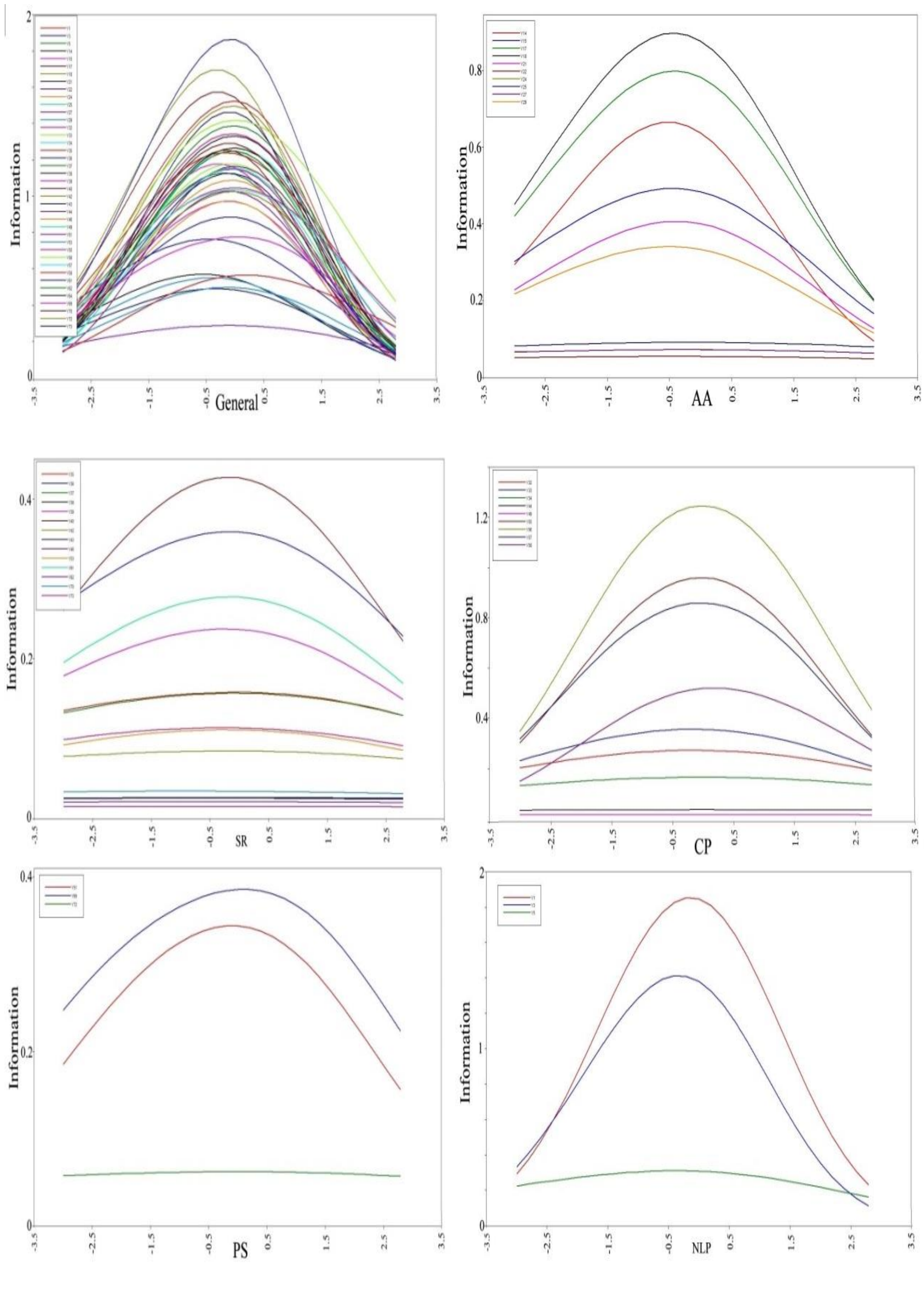

Figure 4. Item information curves

Figure 5 illustrates test information in regard to each factor. Note that because the domain factors were extracted as what was not accounted by the strong general factor, information provided by domain factors were much less than that provided by the general factor. Still, it complemented the general factor and provided substantial meaning in understanding the structure of HBGSI.

Substantial meaning. Among the 40 items selected, one item (Item 64) was free of domain loading because this item had nonsignificant influence by the domain loading in the model. The lack of domain grouping for this item is potentially confusing for practitioners who might be interested to learn and use this instrument in the field. A closer examination of the item indicated that the substantial meaning of nonverbal abilities had already been included in other selected items (Item 69), thus supported the elimination of this item from the final selection (Irby, personal communication, June 5, 2014). Thus we have a final selection of 39 items as shown in Table 7

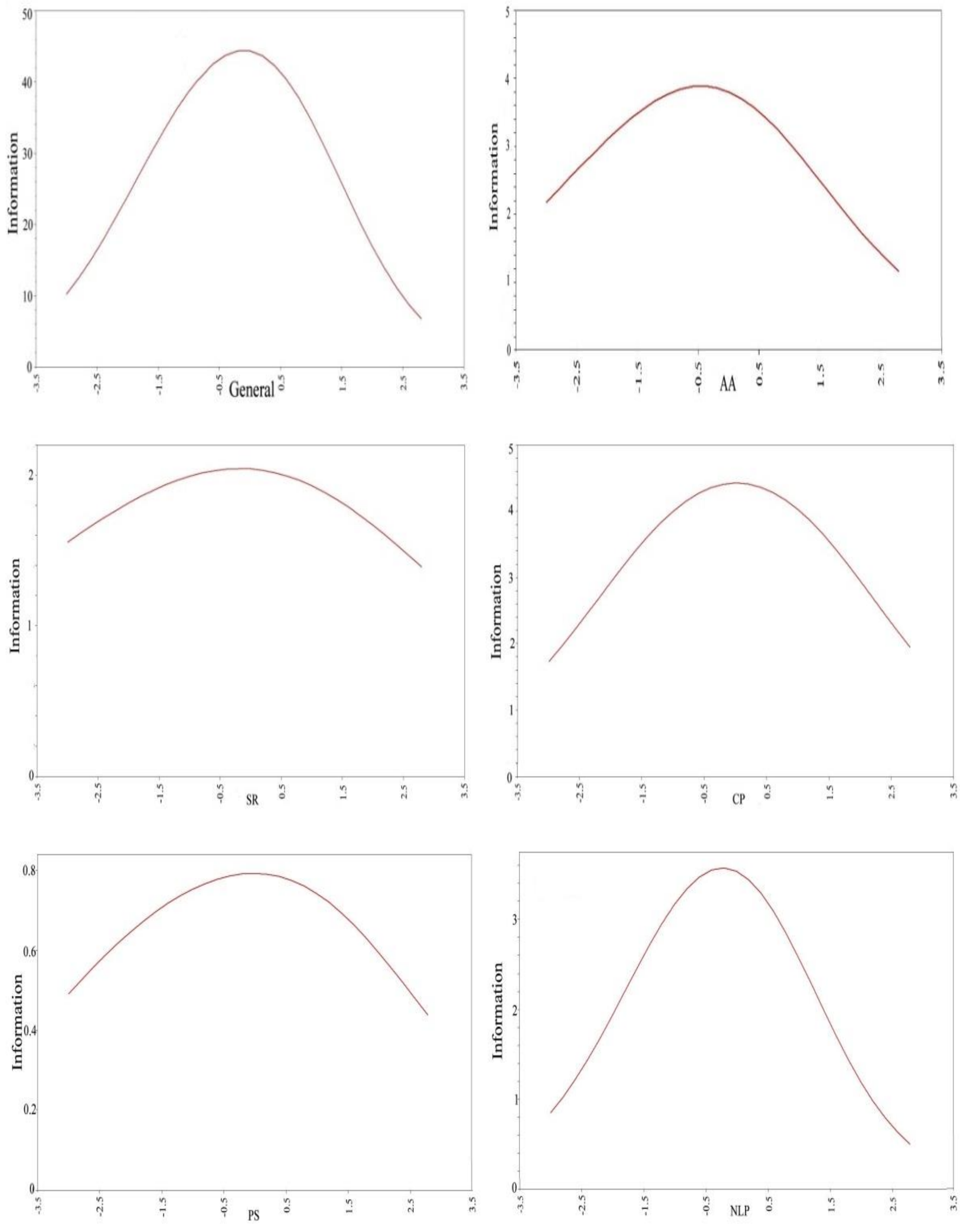

Figure 5. Test information curves

Table 7

Items Selected for HBGSI-S (Short Version of HBGSI)

Item	Description	Domain
Item_01	Likes to read in native language; is a proficient reader in native language	NLP
Item_03	Likes to speak in native language; is a proficient speaker in native language	NLP
Item_05	Aware of own language/culture and takes pride in language/culture	NLP
Item_14	Has strong emotional support from families	SR
Item_15	Values education; sees education as a way to improve status	SR
Item_17	Appears to have sustained motivation to succeed; is persistent	SR
Item_18	Is motivated to learn; exhibits a desire for learning	SR
Item_21	Indirect in giving criticism; avoid conflict; is amenable	SR
Item_22	Possesses leadership qualities in relation to working in the peer group; works well with others	SR
Item_24	Demonstrates ability for giving advice and judgments in disputes and in planning strategies	SR
Item_25	Effective at setting goals	SR
Item_27	Is able to evaluate events and people	SR
Item_29	Likes to please; sensitivity to the opinions of others	SR
Item_32	Exhibits language (speaking) rich in imagery	CP
Item_33	Is imaginative in story telling	CP
Item_34	Exhibits language (writing) rich in imagery	CP
Item_35	Has ability to generalize learning to other areas and show relationships among apparently unrelated ideas	AA
Item_36	Has the ability to use stored knowledge to solve problems	AA
Item_37	Reasons by analog or contrast	AA
Item_38	Talents demonstrated through various projects and interests at home or in the community	AA
Item_39	The relationship between learning and language is consistent in the areas of math and science; level of competency is equal in all of those areas	AA
Item_40	Performs at or above grade level in math; has high math abilities; likes to do math problems	AA
Item_42	Perceives cause and effect relationships	AA
Item_43	Is self-directed in activities and is methodological	AA
Item_44	Has an entrepreneurial ability/spirit	CP

Table 7 Continued

Item	Description	Domain
Item_49	Uses intuition	CP
Item_51	Prefers alternative assessments rather than standardized or teacher made tests	PS
Item_53	Needs minimal support in second language acquisition	AA
Item_55	Exhibits creativity in movement, dance and other physical activities	CP
Item_56	His/her performance in arts, music, dance, when using his/her native culture is unique	CP
Item_57	Adept in visual and/or performing arts; is talented in art, music or drama	CP
Item_58	Creative in lyric production to songs; likes to compose poetry	CP
Item_61	Performs at or above grade level in science; likes to do science experiments	AA
Item_62	Appears to have more patience in dealing with tasks not easily resolved	AA
Item_69	Performs better on spatial fluency tasks as opposed to verbal fluency tasks (fluency is designed as the generation of many ideas - may be verbal on nonverbal)	PS
Item_70	Exhibits steadfast self-concept and self-confidence	AA
Item_72	Has the ability to meaningfully manipulate symbolism in his/her culture	PS
Item_73	Has effective test taking skills	AA

Looking back at clusters. Thus we have selected 39 items for the short version of HBGSI (HBGSI-S), as shown in Table 7 above. All clusters were represented in the short version of HBGSI. A breakdown of number of items retained is presented in Table 8. The highest proportions of items were retained in cluster Achievement and Creative Performance while the least proportion was retained in Familial. Factor wise, the five domain factors each have 10(SR), 14(AA), 9(CP), 3(PS), 3(NPL) items, with one additional item (Item 64) on general construct only.

Table 8

Breakdown of Number of Items in Each Cluster in the Original and Short Version

Cluster	Original	Short
Social and Academic Language	4	2
Cultural Sensitivity	3	1
Familial	7	1
Motivation	4	3
Collaboration	13	6
Imagery	3	3
Achievement	15	11
Support	5	2
Creative Performance	5	4
Problem Solving	10	3
Locus of Control	8	3

HBGSI-Short with HBGSI. These 39 items finally selected for the short version of HBGSI well represented the 77 items in the original HBGSI. Total score correlation of the HBGSI-S with HBGSI is .99. Identification accuracies were also high, with slight variations depending on the rate of referral. Using the original HBGSI as benchmark, HBGSI-S had an accuracy of 95.4% if a teacher referred the top half of students for screening, and no false negatives nor false positives were found in the top 37.1%.

Accuracy increased to 95.9% if a teacher referred only the top quartile in the classroom, and an even higher accuracy of 98.0% if a teacher referred only the top 10% in the classroom. A detailed breakdown of classification statistics is listed in Table 9.

Table 9

Classification Statistics

Statistics	Screening Rate		
	50	25	10
Sensitivity	95.3%	92.7%	90.2%
Specificity	95.7%	97.0%	98.9%
Precision	95.6%	91.1%	90.2%
Negative predictive value	95.3%	97.6%	98.9%
Accuracy	95.5%	95.9%	98.0%

Although accuracy was generally high and increased as we our selection criteria got more stringent, on the downside, sensitivity and precision declined as the selection pool narrowed down, suggesting that the shortened version might come at a trade-off when we use it to select the very top of students. Caution is advised but without posing a threat to the integrity of HBGSI-SS as teachers typically recommended a larger proportion (e.g. the top 50%) of students in a classroom when using HBGSI in practice.

Summary

In this section, three research questions were answered and the answers progressively led to a shortened version of HBGSI. Results of the first research question revealed that there was a strong factor in HBGSI, however, multiple factors co-existed. Based on a bifactor model, a strong general factor and five domain factors were found. The general factor was not dominant enough to warrant a unidimensional solution. A multidimensional IRT model was established and provided answers to the second research question. Items with desirable discrimination power and location were then selected to form a shorter version of HBGSI, providing answers to the third research question.

CHAPTER V

DISCUSSIONS AND CONCLUSION

Demographic changes continue in Texas and in the nation with a rapidly growing Hispanic population. It is projected that Hispanics will make up 45% of the nation’s population growth by 2030 (Passel, Cohn, & Lopez, 2011), and 29% of total population in U.S. (Passel & Cohn, 2008). As Hispanic student become the majority group in Texas public schools (TEA, 2014), enrollment of Hispanic students in G/T programs does not reflect the growth patterns of the student population. Equality in education has been a concern for many researchers, as there are an increasing number of unidentified gifted bilingual students.

According to a State Initiative published by TEA (Slocumb & Olenchak, 2006), even though goal has been set as early as 1996 to have the population in gifted/talented program reflect the population of the total district, little progress has been made. There is still a gap between Hispanic student enrollment in gifted programs and in general public school systems. Nationwide, the likelihood for Hispanic students and African American students to be identified for gifted programs is about only half of the likelihood for White students (National Research Council, 2002).

The facts seem to suggest that identification of gifted students in public schools needs a make-over. Culturally and linguistically appropriate measures need to be in place to ensure that students from diverse backgrounds have equal access to gifted programs. This general rule applies to all three different stages: the nomination/teacher referral; the assessment; and the final recommendation for placement. HBGSI contributes to the pool of instrument in the teacher referral stage as an instrument that takes a students' social-linguistic background into consideration, especially for Hispanic Bilingual students.

Interpreting HBGSI: Giftedness in General and in Domains

With theoretical support of the widely recognized g-factor in intelligence and giftedness literature, and the equivalence in transformation between second order models and bifactor models, an exploratory bifactor model was viable. Bifactor models allowed the analysis of a broad construct and permitted conditional dependence within subsets of items that form a number of group factors (Reise, 2012). Results of this study suggested a construct of general giftedness along with five domain constructs: Academic Achievement (AA); Social Responsibility (SR); Creative Performance (CP); Problem Solving (PS) and Native Language Proficiency (NPL).

Factors

The domain of *Academic Achievement* describes abilities related with performing academic tasks. Items in this domain also exhibited strong influence by the general factor. Academic performance is widely believed to have a strong link to intelligence (Lubinski, 2004). Items included ability to reason, generalize, perceive causal relationships; as well as specific performances in Math and Science. The item with the strongest loading/discrimination to this domain is Item 36: Has the ability to use stored knowledge to solve problems. This was particularly interesting because it coincided with the fundamental philosophy of bilingual education that when students learn a second language, they use existing knowledge of first language to learn multiple aspects of the second language: their knowledge and skills transfer across languages. According to Cummins (2000), children with a solid foundation in their first language develop stronger literacy skills in the second language used in schools. Bilingual students do not learn a second language from scratch: they are already acquainted with content concepts and empowered by their literacy skills in first language. In addition, transfer happens in two ways, concepts, language, and literacy skills can transfer from school language to home too (Cummins, 2000). Thus, it was not surprising that a bilingual student who were good at utilizing stored knowledge seemed to know the rules of the game better, were likely to perform well in developing academic language and academic contents, and also were deemed gifted.

The domain of *Social Responsibility* describes the students' awareness of his/her social role: as a student to value education and be motivated to learn; as a family member, to have emotional support from the family; as a social member, to be sensitive to personal relations; and as a Hispanic descendant, to be respectful and amenable as the culture obliges. In contrast to academic performance, social emotional aspects of a student are often inadequately addressed in the process of giftedness identification. Traditional intelligence tests tend to emphasize cognitive and analytical skills, and neglect to test a students' social and emotional development (e.g., Naglieri, 2014). The simplistic trend to exclude "social emotional" aspect of a person's ability needs to be adjusted, as emotional coping and social skills are just as important, if not more, for long term success of any student. To ignore these abilities would be narrow-minded and inadequate in preparing a student for the world outside classrooms.

Cultural values are most reflected in items in this domain of *Social Responsibility*. Such items described characteristics such as being indirect in giving criticism, avoiding conflict, liking to please and being sensitive to others' opinions, and valuing education, etc. One example is item 12: Respects all authority figure, is respectful. This is a trait typical of the Hispanic culture but possibly not emphasized strongly in the mainstream culture, where instead students are encouraged to challenge authorities. Based on this item's thresholds in the MIRT model, it was so common among the Hispanic bilingual students in this sample, that average students were more likely than not to obtain full score for this item.

The domain *Creative Performance* evaluates students' performances in various activities: music, visual and performing art, poetry, physical activities, games, etc. Creativity involves connecting knowledge and synthesizes, innovate, and invent novelty. It has long been recognized that the highest test takers might score lower in creativity (Renzulli, 1977). Researchers have regarded creativity and intelligence to be two entities only modestly or not related (e.g. review by Batey and Furnham, 2006). Still, debates exist: Nusbaum and Silvia (2010) argued that certain aspects of cognition are central to creative thoughts. In my analysis, the finding of the domain creative performance after the orthogonization seemed to support that creativity are at least an entity different enough from academic performance, and is worth attention of its own. Creativity shines through all channels, and needs to be cultivated so too. Results seemed to indicate that compared to other artistic and physical activities, it was most difficult to be creative in poetry, for the 1^{st} to 3^{rd} graders in this study. It could be speculated that more advanced literacy and language skills for poetry had not been well developed by young children. Although young children encounter poems and poem-like words and phrases in classroom and media (Zeece, 1998), they were more receptive rather than encouraged to be productive in poetic experiences. Few adults share poetry with young children in a formal setting (Glazer, 1997) thus students possibly lack ample opportunities creating poetic work especially when their language skills were deemed insufficient. Results seemed to suggest that it was easier to observe and rate creative performance in physical activities and visual and performing arts. It is thus suggested that abundant varieties of activities be arranged for students inside and outside classrooms to expose them to, and engage them in constructive activities.

The domain *Problem Solving* concerns unique characteristics associated with students' preference to approach a problem and find a solution: learning activities in a group rather than individually, global thinking rather than analytical, alternative assessment. In general, items in this domain were less discriminant in regard to the construct of general giftedness. Researchers cautioned against treating styles simplistically and stereotypically label students based on their ethnicity, gender, or SES status, but rather, learning style should be understood in social and historical context and inform diverse teaching and learning strategies (Gutiérrez & Rogoff, 2003). Characteristics included in this domain concerns with the approaches students take to solve a problem and the flexibility and resourcefulness in their approaches as called for by different contexts.

The domain *Native Language Proficiency* included items specifically related to linguistic skills in the native language, including the four aspects of linguistic: listening, speaking, reading and writing. Results seemed to suggest that for grade 1 to grade 3 students, speaking and reading skills are the most discriminant. Listening and Reading are receptive skills and speaking and writing are productive skills. It is worth noting that both types of skills were represented in the top items. Writing skills were less discriminant, possibly because young children are still at the initial phase of developing writing skills and have less variance in that regard. Similarly listening skills might be comparatively better developed as children have most exposure to oral language at home and also in a second language environment, thus less variances were found in listening rather than speaking and reading.

Finally it was worth noting that complex structure was initially found in the results of the exploratory models: some items were found to have significant loadings on more than one domain. Complex structure is generally not preferable in factor analysis. Forcing items to load on only one domain factor potentially inflated the general factor. However, considering that domain loadings were much smaller compared to general loading in the exploratory model, and all items were needed in MIRT model to preserve the integrity of the instrument, it was decided that domain factor item would be the one with higher loadings. Model fit potentially increases if we relax the constraint of a restricted bifactor structure.

Cluster vs. Bifactor Structure

Each of these five domain factors is composed of a collection of items originally grouped under different clusters. All clusters were found to have contributed to more than one domain factors, in other words, there were no such clusters where all the items were found to all belong to the same domain. A difference in the modeling method possibly contributed to the difference in the cluster structure and bifactor structure. Even though it looked like that the factor structure was not strictly consistent or hierarchical with the cluster structure, it could be argued that there are at least some items that were grouped better according to its face value.

For example, item 41 and item 62 had almost the same characteristics, querying the students' performance and interest in math/math problems and science/science experiments. In the cluster-wise structure, these items were grouped in two respective clusters: Item 41 in achievement and Item 62 in problem solving. The separation of two structurally similar and substantially close-knit items in two separate clusters could potentially confuse teachers. The cloud was lifted in the bifactor model as both items were unified in the same domain: Academic Achievement. For another example, in the bifactor structure, Items 59 and 67 were grouped in one domain as they both concern performances in groups: whether their creativity was found in group activities or whether their identity is observed in relation to groups. Both items focused on the contrast between group and individual, and it did make sense that these items were found to be in the same domain even though they were originally in different clusters. In a similar fashion, Item 45 and 54, although clustered respectively in Achievement and

Support, both centered on students' native language skills---vocabulary and working command. Thus grouping them in the same domain of NLP aligned them with other linguistically related domain.

In summary, through the bifactor IRT model, we gain insights into the structure of HBGSI. There was a strong construct of general giftedness that influenced all items in HBGSI to different extent, and the remaining variances can be explained by five domains that each complements the general giftedness construct.

HBGSI-S: The Shortened HBGSI

Adaptation of HBGSI was based on a multidimensional IRT model, with a sample of first, second and third graders in a big urban districts in Texas. The main purpose of the adaptation is to identify items that best reflect giftedness in Hispanic bilingual students and provide an optional shortened version of HBGSI for ease of administration.

For the short version of HBGSI, items that were found to be not discriminate among gifted students were excluded: for example, attending school regularly (Item 16), completing homework (Item 55), or performing better when teacher expresses confidence in his/her ability (Item 51). These characteristics were found in gifted students, but also frequently found in average or even below average students. In terms of item scores, a majority of students would have scored high on these items, making the item less effective to be used for giftedness identification.

Items where the hidden constructs were better manifested in other items were also eliminated. For example, Item 47 "has a rich sense of humor, is motivated by humor” is eliminated because it was in the lower 50 percentile for both general giftedness and its own domain: creative performance. This does not mean that this item, if singled out, is a bad item to tell gifted students apart from average. In fact, understanding and appreciating humor is an important trait of giftedness----it requires an understanding of the ‘serious’ subject matter as well as a novel way to reshape it with a lighter tone. The elimination only meant that in this model, the ability to understand and the ability to create was better represented by other items, making the inclusion of this item being less meaningful. This was also likely sample specific---young children might not be able to appreciate humor as well as older children or adults do, thus making this item less discriminating in the model where the samples are taken from a younger population. This item might be a very good one if measured against a sample of adults.

HBGSI was developed with cultural-social context in mind (Irby & Lara-Alecio, 1996). Even though only one third of the original three items in the Cultural cluster were retained in the short version, cultural and social influences can be found in other items throughout the shortened instrument. For example, problem solving were influenced and shaped by cultural socialization and multiple aspects of cultures as operationalized by countries (Joy & Kolb, 2009). Many items retained in the short version reflected unique characteristics of the Hispanic culture. Items in social responsibility reflected cultural values such as Valuing education (Item 15), Avoiding conflict (Item 21), Liking to please and being sensitive to others (Item 29). However, some strongly cultural related characteristic, such as item 12 as described in the previous section, were so commonly found and ‘always exhibited’ by many students, that this item become less discriminant for gifted screening purposes. It is for that reason that this items was not retained in the short version even though it carried strong cultural-social influences.

Criterion validity was not provided for the shortened version of HBGSI, because in HBGSI, the fundamental definition of giftedness differed from what was used in standardized tests. HBGSI takes consideration of social-linguistic context, in addition to above average ability, motivation and creativity (Irby & Lara-Alecio, 1996), while most standardized tests are concerned about a smaller set of human abilities, such as verbal skills, or nonverbal skills.

A further distinction should be made that HBGSI-S measures *typical* performance while standardized tests measures *maximum* performance. Intelligence-as-typical performance (Ackerman, 1997) relates to personality, vocational interests besides intelligence in the traditional sense. Four facets of social, clerical/conventional, science/math, and intellectual cultural were identified for measuring intelligence as typical performance (Ackerman, 1997). Therefore, the significance to correlate HBGSI to a standardized test for criterion validity is limited. If needed, items that are most influenced by general factor and “academic achievement” can be pulled, to compose a version more predictable for results of standardized tests.

Pluralism for definition of giftedness was recommended by TEA in the State Initiative where the broadest defensible definition of gifted/talented was suggested (Slocumb, & Olenchak, 2006). Thus the short version of HBGSI was adapted with the inclusive purpose in mind to keep HBGSI as comprehensive as possible. The high correlation of total score of HBGSI-S and HBGSI seemed to suggest that HBGSI did retain the many aspects that the original HBGSI aimed to model.

Implications for Practice

It has long been recognized that identification of giftedness is a complex process (Renzulli, 1970) and is even more complicated with more cultural and social influences and the influx of demographic changes. Relying on standardized tests only is simplistic. Educators need to make the identification process and gifted programs themselves more inclusive, both in the sense of its student demographics, and also in the conceived notation of giftedness.

No measures should be taken blindly as a gold standard, and used as a single (NRC, 2002). A collective set of non-verbal tests, observation instruments, portfolios, performance projects, extensive interviews tests were advised by researchers to be used for gift identification, along with professional training opportunities for teachers to become more culturally competent in today's multicultural school settings, had been recommended (Ramos, 2010). Furthermore, measures needs to be adopted based on the demographics of different districts and schools (Slocumb & Olenchak, 2006). The set of instrument recommended for use in a school with a majority of Hispanic Students who are ELLs and coming from low SES families, should be very different from another set recommended to a school with a majority of White Middle class students. A local norm is critical when dealing with heterogeneous population.

Teacher referral is usually the first step in gifted identification (Elhoweris, Mutua, & Alsheikh, 2005) and inadequate teacher referrals, tests, and policies and procedures contribute greatly to segregation in gifted classes (Ford, 2012). HBGSI-S could be used in the initial phase of screening, especially for Hispanic low-SES English Language Learners (Fultz, Lara-Alecio, Irby & Tong, 2013).

With the best intent to understand LCD students, teachers are not always familiar with how to find the gifted child, i.e., what characteristics to look for. Traditional teacher education has been reported to be weak on preparing teachers from mainstream to understand the needs of LCD students (Ford, 2003), but it will take decades before the teaching force catches up with the profile of the general student population. Misconceptions still exist such as a student needing to be fluent in English to be gifted. Teachers might overlook talent in Hispanic bilingual students because their needs for cognitive development were disguised under their still-developing linguistic proficiency in English. On the other hand, a teacher might overlook the needs of linguistic support for a *smart kid*, ignoring the fact even a gifted Hispanic, who maybe quick to learn, can still benefit from strategies such as sheltered instructions to develop both CALP and content knowledge.

Besides cultural and linguistic diversity, identification of Hispanic Bilingual students is also confounded by other factors. As of 2011, as high as 37% of Hispanic population in Texas, 17 years old or younger, are in poverty (Pew Research, n.d.). Many Hispanic bilingual students come from families with low Social Economic Status, which influenced their behavior in school. It has been reported that students from poverty often manifest their giftedness in negative ways (Slocumb & Olenchak, 2006). For example, an inquisitive mind might raise questions that challenge the teacher or remarks with "shock value", which can be misunderstood by teachers that grew up and were trained by middle class standards (Slocumb & Olenchak, 2006), and misinterpreted coupled with the student's ethnicity. Therefore, throughout the process of gifted student identification, student behaviors need to be interpreted within context.

HBGSI can not only help teachers in the referral process, but also aid the teacher in capturing characteristics that gifted and potentially gifted students exhibit in their daily activity. Teachers can refer to items in HBGSI for evidence of giftedness in the students' behaviors, and also create classroom opportunities for students to showcase their talent.

More important than identifying gifted students, is cultivating talent. It is important that gifted students are provided with instructions and opportunities that are at the level suitable for their developmental stage. Programs such as bilingual G/T programs and two-way bilingual G/T programs have been proposed by scholars (Bernal, 2002). Activities fostering giftedness and talents are feasible in and outside a gifted program. If a student exhibit talent in music, the teacher can nurture this talent by encouraging for performances in school plays, incorporating music into daily instructions, connecting musical events in school and community to the classroom, so on and on. Giftedness is in the eyes of the beholder and HBGSI provide a lens to help the teachers see and understand gifted Hispanic bilingual children.

Limitations

This study based on data collected from ELL student in an urban district in Texas, and thus is limited by the sample characteristics used in the study. In an extreme case, in IRT, if a highly discriminant item was administered to a group of people with identical latent trait, it is likely that this item will be found to have little discriminant power. The small value is due to the monotone sample and not an intrinsic property of the item. Thus the result of this study should be generalized with caution: when used in another district or with a different population, some items might not exhibit high discriminating power as it does in this study. Similarly, items that were not retained in the short version might be very good items for another population, say, Hispanic bilingual adults or professionals.

Another limitation of the study is that item level was only available by itself, and no identification information was available. Thus it was impossible to associate item score to other roster information, such as age, gender, language proficiency, and the lack of information has prevented the possibility of DIF analysis.

Future Work

With the limitations in mind, it is recommended that future studies be extended to more students who are Hispanic and bilingual. It would be most beneficial if cross validation can be performed on another group of students with similar cultural and linguistic backgrounds.

Because a broader definition of giftedness is adopted in HBGSI, comparing its scores to standardized tests would be unfair. On the other hand, valuable insights could be gained if follow up studies could be conducted with students who were identified as HBGSI, to track their academic and career success in the long term. To understand the comprehensive measure of gift identification, and more importantly, to identify elements that contribute to students' overall success and foster those elements, causal analysis could be ventured. It is worthwhile to explore what characteristics were underlying explanations and possibly the basis of other exhibits of giftedness.

REFERENCES

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*,*121*,219-245.

Ackerman, P. L. (2009). Personality and intelligence. In P. J. Corr & G. Matthews (Eds.), *The Cambridge handbook of personality psychology*, (pp. 162-174). New York: Cambridge University Press.

American Educational Research Association, American Psychological Association, National Council of Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: Author.

Agrawal, N., Sinha, S. N., & Jensen, A. R. (1984). Effects of inbreeding on Raven Matrices. *Behavior Genetics*, *14*, 579-585.

Anastasi, A. (1992). What counselors should know about the use and interpretation of psychological tests. *Journal of Counseling and Development*, *70* (5), 610-615.

Baker, F. B., & Kim, S. (2004). *Item response theory: Parameter estimation techniques.* New York: Marcel Dekker, Inc.

Banks, J. A. & Banks, C. M. (Eds.) (1995). *Multicultural education: Issues and perspectives*. Boston: Allyn and Bacon.

Bann, C. M., Kobau, R., Lewis, M. A., Zack, M. M., Luncheon, C., & Thompson, W. W. (2012). Development and psychometric evaluation of the public health surveillance well-being scale. *Quality of Life Research*, *21*(6), 1031-1043.

Bartholomew, D. J. (1983). Latent variable models for ordered categorical data. *Journal of Econometrics*, *22*, 229-243.

Bartholomew, D. J. (1985). *A unified view of factor analysis, latent structure analysis and scaling.* Invited paper presented at the 4th European Meeting of the Psychometric Society and Classification Societies, Cambridge.

Batey, M., & Furnham, A. (2006). Creativity, intelligence, and personality: a critical review of the scattered literature. *Genetic, Social, and General Psychology Monographs*, *132*(4), 355−429.

Becker, K. A. (2003). *History of the Stanford-Binet intelligence scales: Content and psychometrics.* Itasca, IL: Riverside Publishing.

Bernal, E. (2002). Three ways to achieve a more equitable representation of culturally and linguistically different students in GT programs. *Roeper Review*, *24*, 82-88.

Biderman, M. (2013). *Applications of bifactor models to Big Five data*. Presented at 28th Annual Conference of The Society for Industrial and Organizational Psychology, Houston, TX.

Binet. A., & Simon, T. (1916). *The development of intelligence in children.* (E.S. Kites, Trans.) Baltimore: Williams & Wilkins. Retrieved from: http://books.google.com/books?id=jEQSAAAAYAAJ&dq=The%20development%20of%20intelligence%20in%20children%20Binet&pg=PA1#v=onepage&q&f=false

Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full information item factor analysis. *Applied Psychological Measurement*, *1988*(12), 261-280.

Brown, A., & Lopez, M. H. (2013). *Mapping the Latino population, by State, County and City*. Retrieved from http://www.pewhispanic.org/2013/08/29/mapping-the-latino-population-by-state-county-and-city/

Callahan, C. M. (2005). Identifying gifted students from underrepresented populations. *Theory into Practice, 44*(2), 98-104.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge, UK: Cambridge University Press.

Castellano, J. A. (2011). Hispanic students and gifted education: New outlooks, perspectives, and paradigms. In Castellano, J. A. & Frazier, A. D. (Eds.), *Special populations in gifted education* (pp. 249-269). Waco, TX: Prufrock Press Inc.

Chamorro-Premuzic, T., Moutafi, J. & Furnham, A. (2005). The relationship between personality traits, subjectively-assessed and fluid intelligence. *Personality and Individual Differences*, *38*(7), pp. 1517-1528.

Ceci, S. J. & Williams, W. M. (1997). Schooling, intelligence, and income. *American Psychologist*, *52*, 1051-1058.

Cohen, L. M. (1988). Meeting the needs of gifted and talented minority language students: Issues and practices. *National Clearinghouse for Bilingual Education*, *8*, 2-9.

Contreras-Vanegas, A. L. (2011). *The psychometric properties of the Hispanic Bilingual Gifted Screening Instrument (HBGSI).* (Ph.D. dissertation, Texas A&M University). Retrieved from http://library.tamu.edu./about/thesis-dissertations

Contreras-Vanegas, A. L., Lara-Alecio, R., Tong, F., Irby, B. J., & Pollard-Durodola, S. (2012). The inter-rater reliability of the Hispanic Bilingual Gifted Screening Instrument (HBGSI). *National Forum of Applied Educational Research Journal*, *25*(3), 1-13.

Cummins, J. (2000). *Language, power and pedgogy: Bilingual children in the crossfire.* Clevedon, U.K.: Multilingual Matters.

Dai, D. Y. & Renzulli, J. S. (2008). Snowflakes, living systems, and the mystery of giftedness. *Gifted Child Quarterly*, *52*, 114-130.

Das, J. P., Naglieri, J. A., & Kirby, J. R. (1994). *The assessment of cognitive processes: The PASS theory of intelligence.* Boston: Allyn and Bacon.

DeYoung, C. (2011). Intelligence and personality. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 711–737). New York: Cambridge University Press.

du Toit, M. (ed.) (2003). *IRT from SSI: BILOG-MG MULTILOG PARSCALE TESTFACT*. Lincolnwood, IL: Scientific Software International, Inc.

Ebesutani, C., Drescher, C.F., Reise, S. P., Heiden, L., Hight, T., Damon J. & Young, J. (2012): The Loneliness Questionnaire–short version: An evaluation of reverse-worded and non-reverse-worded items via Item Response Theory. *Journal of Personality Assessment*, *94*(4), 427-437.

Elhoweris, H., Mutua, K., & Alsheikh, N. (2005). Effect of children's ethnicity on teacher's referral and recommendation decisions in gifted and talented programs. *Remedial and Special Education*, *26*, 25-31.

Embretson, S. E., & Hershberger, S. L. (Eds.). (1999). *The new rules of measurement: What every psychologist and educator should know*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Embretson, S. E. & Reise, S. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum Publishers.

Esquierdo, J. J. (2006). *Early identification of Hispanic English language learners for gifted and talented programs*. (Ph.D. dissertation, Texas A&M University). Retrieved from http://library.tamu.edu/about/thesis-dissertations

Esquierdo, J. J. & Arreguín-Anderson, M. (2012). The "invisible" gifted and talented bilingual students: A current report on enrollment in GT programs. *Journal for the Education of the Gifted*, *35*(1), 34-47.

Erwin, J., & Worrell, F. C. (2012). Assessment practices and the underrepresentation of minority students in gifted and talented education. *Journal of Psychoeducational Assessment*, *30*(1), 74-87.

Fagan, J. R. (2000). A theory of intelligence as processing: Implications for society. *Psychology*, *Public Policy*, *and Law*, *6*,168.179.

Ford, D. (2012) Culturally different students in special education: Looking backward to move forward. *Council for Exceptional Children*, *78*(4), 391-405.

Ford, D. Y., & Grantham, T. C. (2003). Providing access for culturally diverse gifted students: From deficit to dynamic thinking. *Theory into Practice*, *42*(3), 217-225.

Ford, D.Y., Grantham, T.C., & Whiting, G.W. (2008). Culturally and linguistically diverse students in gifted education: Recruitment and retention issues. *Council for Exceptional Children, 74*(3), 289-306.

Fox, J. P. (2010). *Bayesian item response modeling: Theory and applications*. New York: Springer.

Fultz, M. (2004). *Psychometric validation of the Hispanic Bilingual Gifted Screening Instrument (HBGSI)* (Ph.D. dissertation, Texas A&M University). Retrieved from http://library.tamu.edu./about/thesis-dissertations

Fultz, M., Lara-Alecio, R., Irby, B. J., & Tong, F. (2013). The Hispanic Bilingual Gifted Screening Instrument: A validation study. *National Forum of Multicultural Issues Journal*, *10*(1), 1-26.

Galton, F. (1869). *Hereditary genius: Its laws and consequences*. London: MacMillian and Co.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Glazer, J. (1997). *Introduction to children's literature*. Upper Saddle River, NJ: Merrill.

Gonzalez, V., & Yawkey, T. (1993). The assessment of culturally and linguistically different students: Celebrating change. *Education Horizons*, *72*(1), 41-49.

Gottfredson, L. (2004). Realities in desegregating gifted education. In Boothe, D. & Stanley J. C., *In the eyes of the beholder: Critical issues for diversity in gifted education* (p. 139-155). Waco, TX: Prufrock Press.

Guilford, J. P. (1950). Creativity. *American Psychologist*, *5*(9), 444–454.

Gutiérrez, K. D. & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher, 32*(5), 19-25.

Gustafsson, J., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, *28*, 407–434.

Harris, B., Plucker, J. A., Rapp, K. E., & Martinez R. S. (2009). Identifying gifted and talented English language learners: A case study. *Journal for the Education of the Gifted*, *32*(3), 368-393.

Harris, B., Rapp, K. E., Martinez, R. S., & Plucker, J. A. (2007). Identifying English language learners for gifted and talented programs: Current practices and recommendations for improvement. *Roeper Review*, *29*(5), 26-29.

Hartley, E. A. (1987). *How can we meet all their needs? Incorporating education for the gifted and talented into the multicultural classroom*. ERIC Document Reproduction Service No. ED336968.

Horowitz, F. D. (2000). Child development and the PITS: simple questions, complex answers, and developmental theory. *Child Development*, *71*, 1-10.

Irby, B., Hernandez, L., Torres, D., & Gonzalez, C. (1997). *The correlation between teacher perceptions of giftedness and the Hispanic Bilingual Screening Instrument*. Doctoral dissertation, Sam Houston State University, Huntsville, Texas.

Irby, B. J., & Lara-Alecio, R. (2003). *The Hispanic Bilingual Gifted Screening Instrument: A factor analysis report*. Retrieved from www.teachbilingual.com

Irby, B. J., Lara-Alecio, R., Mathes, P.G., Rodriguez, L., & Guerrero, C. (2008). *Promoting bilingualism and biliteracy: Programmatic difference between one-way dual language and transitional bilingual programs*. Paper presented at the annual meeting of National Association of Bilingual Education, Tampa, FL.

Irby, B. J., Lara-Alecio, R., Quiros, A. M., Mathes, P. G., & Rodriguez, L. (2004). *English Language and Literacy Acquisition evaluation research program (Project ELLA): Second annual evaluation report.* Washington, DC: Institute for Educational Sciences, U.S. Department of Education.

Irby, B. J., Tong, F., Lara-Alecio, R., Mathes, P. G., Acosta, S., & Guerro, C. (2010). Quality of instruction, language of instruction, and Spanish-speaking English language learners' performance on a State reading achievement test. *TABE Journal*, *12*(1), 1-42.

Irby, B.J., Tong, F., Polnick, B., Lara-Alecio, R., & Fan., Y. (2013). Early identification of English language learners for gifted education programs using the Hispanic bilingual gifted screening instrument. *He Kupu eJournal*, *3*(3). Retrieved from: http://www.hekupu.ac.nz/index.php?type=issue&issue=18

Jensen, A. R. (1986). g: artifact or reality? *Journal of Vocational Behavior*, *29*, 301-331.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CN: Praeger.

Jensen, A. R., & Reynolds, C. R. (1982). Race, social class and ability differences on the WISC-R. *Personality and Individual Differences*, *3*, 423-438.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*,141-151.

Lakeshore. (1997). *Question of the day*. Carson, CA: Lakeshore Learning Materials.

Lara-Alecio, R., & Irby, B.J. (March, 1993). *Identifying the bilingual gifted*. Paper presented to the 22nd National Association of Bilingual Education Annual Conference, Houston, TX.

Lara-Alecio, R., Irby, B., Tong, F., & Mathes, G. P. (2009). *English Language and Literacy Acquisition (Project ELLA):Final and fifth year evaluation performance report*. Washington, D.C.: U.S. Department of Education.

de Leeuw, J. (1983). Models and methods for the analysis of correlation coefficients. *Journal of Econometrics, 22*, 113-137.

Lohman, D. F. & Rocklin, T. (1995). Current and recurrent issues in the assessment of intelligence and personality. In D. H. Saklofske & M. Zeidner (Eds.), *International handbook of personality and intelligence* (pp. 447-474). New York: Plenum.

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "general intelligence, objectively determined and measured*." Journal of Personality and Social Psychology, 86*, 96-111.

Masters, G.N. (1982) A Rasch model for partial credit scoring. *Psychometrika*, *47*,149–174.

McKenzie, J. A. (1986). The influence of identification practices, race, and SES on the identification of gifted students. *Gifted Child Quarterly, 30*, 93-95.

McGuire, F. (1994). Army alpha and beta tests of intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of intelligence* (pp. 125-129). New York: Macmillan.

Muñoz-Sandoval, A., Cummins, J., Alvarado, G., & Ruef, M. (1998). *Bilingual Verbal Ability Tests: Comprehensive manual*. Itasca, IL: Riverside Publishing.

Most, B. & Zeidner, M. (1995). Constructing personality and intelligence test instruments: Methods and issues. In D. Saklofske & M. Zeidner (Eds.). *International handbook of personality and intelligence* (pp. 475- 503). New York: Plenum.

Muthén, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, *22*, 43-65.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*, 115-132.

Muthén, B., & Asparouhov. T (2012). *New developments in Mplus* version 7: Part 1. Retrieved from http://www.statmodel.com/download/handouts/MuthenV7Part1.pdf

Muthén, L.K. & Muthén, B. O. (1998-2012). *Mplus User's Guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

National Association for Gifted Children. (2008). *The role of assessment in the identification of gifted students*. Retrieved from http://www.nagc.org/uploadedFiles/assessment%20pos%20paper%20final.pdf

National Research Council. (2002). *Minority students in special and gifted education*. Washington, DC: The National Academies Press.

Naglieri, J. A. (1997). *Naglieri nonverbal ability test*. San Antonio: The Psychological Corporation.

Naglieri, J. A. (April 10, 2014). *Cognitive Assessment System-II*. (CAS-2). Presented via Skype in Invited Talk in Department of Educational Psychology. College Station, TX: Texas A&M University.

Naglieri, J. A., & Bornstein, B. T. (2003). Intelligence and achievement: Just how correlated are they? *Journal of Psychoeducational Assessment*, *21*, 244–260.

Naglieri, J. A., & Das, J. P. (1997). *Cognitive assessment system*. Itasca, IL: Riverside Publishing.

Naglieri, J. A., & Ford, D. Y. (2003). Addressing under-representation of gifted minority children using the Naglieri Nonverbal Ability Test (NNAT). *Gifted Child Quarterly*, *47*, 155–160.

Naglieri, J. A., & Jensen, A. R. (1987). Comparison of black-white differences on the WISC-R and the K-ABC: Spearman's hypothesis. *Intelligence*, *11*(1), 21-43.

Naglieri, J., Rojahn J. & Matto, H. (2007) Hispanic and non-Hispanic children's performance on PASS cognitive processes and achievement. *Intelligence*, *35*, 568–579.

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*, 77-101.

Ogbu J. U., & Davis A. (2003). *Black American students in an affluent suburb: A study of academic disengagement*. Mahwah, NJ: Erlbaum.

Ogbu, J. U. (1992). Understanding cultural diversity and learning. *Educational Researcher*, *21*, 5-13.

Okagaki, L., & Sternberg, R. J. (1994). Perspectives on kindergarten: Rafael, Vanessa, and Jamlien go to school. *Childhood Education*, *71*(1), 14-19.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers*, *32*, 396-402.

Ostini, R., & Nering, M. (2006). *Polytomous item response theory models*. New York: Routledge.

Ouyang, M., & Conoley, J. C. (2007). Consultation for gifted Hispanic students: 21st-century public school practice. *Journal of Educational and Psychology Consultation*, *17*(4), 297-314.

Padron, Y. N., & Waxman, H. C. (Eds.). (1999). *Effective instructional practices for English language learners*. Berkeley, CA: McCutchan.

Pew Research. (n.d.). *Hispanic trend project*. Retrieved from http://www.pewhispanic.org/states/state/tx/

Perrone, P. A. & Alernan, N. (1983). Educating the talented child in a pluralistic society. In D. R. Omark & J. G. Erickson (Eds.). *The bilingual exceptional child* (pp. 269-283). San Diego, CA: College-Hill Press.

Plata, M., Masten, W. G., & Trusty, J. (1999). Teachers' perception and nomination of fifth-grade Hispanic and Anglo students. *Journal of Research and Development in Education*, *32*(2), 113-123.

Raven, J., Raven, J. C. & Court, J. H. (2000). *Manual for Raven's progressive matrices and vocabulary scales. Section 3: The standard progressive matrices*. San Antonio, TX: Harcourt Assessment.

Ramos, E. (2010). Let us in: Latino underrepresentation in gifted and talented programs. *Journal of Cultural Diversity*, *17*(4), 151-153.

Reckase, M.D. (2009). *Multidimensional item response theory*. New York: Springer.

Reise, S. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667-696.

Reise, S., & Henson, J.M. (2003). A discussion of modern versus traditional psychometrics as applied to personality assessments. *Journal of Personality Assessment*, *81*(2), 93-103.

Reise, S., Horan, W. P., & Blanchard, J.J. (2011). The challenges of fitting an Item Response Theory model to the Social Anhedonia Scale. *Journal of Personality Assessment*, *93*(3), 213-224.

Reise, S. P., Moore, T. M., Haviland, M. G. (2010). Bifactor models and rotations: Explore the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, *92*(6), 544-559.

Reise, S., Morizot, J., & Hays, R.D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, *16*, 19-31.

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology, 5*, 27-48.

Renzulli, J. S. (1977). *The enrichment triad model: A guide for developing defensible programs for the gifted and talented*. Mansfield Center, CT: Creative Learning Press.

Renzulli, J. S. (1999). What is this thing called giftedness, and how do we develop it? A twenty five year perspective. *Journal for the Education of the Gifted*, *23*, 3-54.

Roid, G. & Barram, A. (2004). *Essentials of Stanford-Binet intelligence scales (SB5) assessment*. Indianapolis, IN: Wiley.

Roid, G. (2003*). Stanford-Binet intelligence scale.* Available at: http://www.riversidepublishing.com/products/sb5/

Ryan, C. (2013). *Language use in the United States: 2011* (American Community Survey Reports, ACS-22). Washington, DC: US Census Bureau. Retrieved from http://www.census.gov/prod/2013pubs/acs-22.pdf

Saccuzzo, D. P., Johnson, N. E., & Guertin, T. L. (1994). *Identifying underrepresented disadvantaged gifted and talented children: A multifaceted approach*. San Diego: San Diego State University.

Samajima, F. (1969). Estimation of latent ability using a response pattern of graded score. *Psychometrika Monograph*, *17*.

Saklofske, D. & Zeidner, M. (Eds.). (1995). *International handbook of personality and intelligence*. New York: Plenum.

Schmid, J., & Lieman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22, 53–61.

Shavinina, L. V., & Seeratan, K. L. (2004). Extracognitive phenomena in the intellectual functioning of gifted, creative, and talented individuals. In L. V. Shavinina & M. Ferrari (Eds.), *Beyond knowledge: Extracognitive aspects of developing high ability* (pp. 73-102). Mahwah, NJ: Lawrence Erlbaum.

Simonton, D. K. (1999). Talent and its development: An emergenic and epigenetic model. *Psychological Review*, *3*, 435-347.

Slocumb, P.D., Olenchak, F. R. (2006). *Equity in gifted education: A state initiative.* Austin, TX: Texas Education Agency. Retrieved from http://www.gtequity.org/docs/equity_in_ge.pdf

Spearman, C. (1927). *The abilities of man.* London, U.K.: Macmillan.

Sternberg, R. J. (1985): *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.

Sternberg, R. J. (1996). *Successful intelligence*. New York: Simon & Schuster.

Sternberg, R. J. (2007). Cultural concepts of giftedness. *Roeper Review*, *29*, 160-165.

Sternberg, R. J. & Grigorenko, E. L. (2004). Intelligence and culture: How culture shapes what intelligence means, and the implications for a science of well-being. *Philosophical Transactions of the Royal Society*, *359*, 1427–1434.

Sternberg, R. J., & Kaufman, J. C. (1998). Human abilities. *Annual Review of Psychology*, *49*, 479-502.

Suzuki, L. A., & Valencia, R. R. (1997). Race–ethnicity and measured intelligence. *American Psychologist*, *52*, 1103−1114.

Takane Y., & de Leeuw J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika, 52*, 393-408.

Terman, L. M. (1916). *The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford Revision and Extension of the Binet-Simon Intelligence Scale*. Cambridge, MA: Riverside.

Texas Education Agency. (2003*). Academic excellence indicator system: District Performance.* Retrieved from http://ritter.tea.state.tx.us/perfreport/aeis/2003/index.html

Texas Education Agency (TEA) (2008). *Equity in gifted/talented education.* Retrieved from http://www.gtequity.org/taskforce.php

Texas Education Agency (TEA) (2009). *Texas State Plan for the education of gifted/talented students – revised 2009*. Retrieved from http://www.tea.state.tx.us/index2.aspx?id=6420

Texas Education Agency. (2014*). Enrollment in Texas public schools, 2012-13.* (Document No. GE13 601 06). Austin, TX: Author. Retrieved from http://www.tea.state.tx.us/acctres/enroll_index.html

Thorndike, R. L. (1987). Stability of factor loadings. *Personality and Individual Differences*, *8*(4), 585–586.

U.S. Census Bureau. (2010). *The Hispanic population: 2010*. Retrieved from http://www.census.gov/prod/cen2010/briefs/c2010br-04.pdf

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41*, 321-327.

Ventriglia, L., & González, L. (2000). *Santillana intensive English.* Miami, FL: Santillana USA.

Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore (MD): Williams & Witkins.

Wechsler, D. (1949*). Wechsler Intelligence Scale for Children*. New York: The Psychological Corporation.

Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale*. New York: The Psychological Corporation.

Wechsler, D. (1974). *Wechsler Intelligence Scale for Children-Revised*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children–III* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale–III*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children–IV*. San Antonio, TX: Harcourt Assessment, Inc.

Wechesler, D., & Naglieri, J. A. (2006). *Wechsler Nonverbal Scale of Ability*. San Antonio, TX: Harcourt Assessment.

Winner, E. (1996). *Gifted children: Myths and realities*. New York: Basic Books.

Wolfram Demonstration Project (2014). *Comparing the normal ogive and logistic item characteristic curves*. Retrieved from http://demonstrations.wolfram.com/ComparingTheNormalOgiveAndLogisticItemCharacteristicCurves/

Worrell, F. C. (2009). Myth 4: A single test score or indicator tells us all we need to know about giftedness. *Gifted Child Quarterly, 53*, 242-244.

Yang, S. & Sternberg, R. J. (1997). Taiwanese Chinese people's conceptions of intelligence. *Intelligence*, *25*(1), 21-36.

Yung, Y. F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika*, *64*, 113–128.

Zeece, P. D. (1998). Dancing words: poetry for young children. *Early Childhood Education Journal*, *26*(2), 101-106.

Ziegler, A. (2005) The actiotope model of giftedness. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (2nd ed., pp. 411-436). Cambridge, UK: Cambridge University Press.

Zwick, W.R. & Velicer, W.F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*, 432-442.